Mikel Martínez-Areta (ed.)

Basque and Proto-Basque

Language-Internal and Typological Approaches to Linguistic Reconstruction

Mikroglottika

Minority Language Studies

Volume 5

PETER LANG
EDITION

Basque and Proto-Basque

Mikroglottika

Minority Language Studies

Volume 5

PETER LANG
EDITION

Mikel Martínez-Areta (ed.)

Basque and Proto-Basque

Language-Internal and Typological Approaches to Linguistic Reconstruction

PETER LANG
EDITION

Bibliographic Information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at http://dnb.d-nb.de.

This Publication was sponsored by

E-ISBN 978-3-653-02701-3 (E-Book)
DOI 10.3726/978-3-653-02701-3
ISSN 1867-190X
ISBN 978-3-631-62649-8

© Peter Lang GmbH
Internationaler Verlag der Wissenschaften
Frankfurt am Main 2013
All rights reserved.
Peter Lang Edition is an Imprint of Peter Lang GmbH

All parts of this publication are protected by copyright. Any utilisation outside the strict limits of the copyright law, without the permission of the publisher, is forbidden and liable to prosecution. This applies in particular to reproductions, translations, microfilming, and storage and processing in electronic retrieval systems.

www.peterlang.de

Hitzaurrea

Blanca Urgell
(Kultura Sailburua, Eusko Jaurlaritza)

Euskara eta haren historia ikertzen aurrerapauso ikusgarriak izan dira azken hamarkadetan. Euskararen bilakaera inoiz baino hobeto ulertzen dugu orain, eta hizkuntza bera osatzen dituzten egiturak nondik eta nola sortu diren zehazteko hipotesi sendo bezain emankorrak garatu ditu euskalaritzak. Ikerketa metodoak eta tresnak hobetuz eta homologatuz joan dira hizkuntzaren iraganetik ahalik eta informazio aberatsena eskuratu nahirik. Koldo Mitxelenak aldarrikatu zuen, bestela maisutzat zeukan Antoine Meilletek esandakoaren aurka, euskara ez dela inola ere historiarik gabeko hizkuntza bat, haren isolamendu genetikoak (ez baitu akitanieraz beste ahaide ezagunik) ez dituela bilakaera historikoa azaltzeko bideak agortzen. Aitzitik, beste hizkuntzetatik hartu dituen maileguetan eta barneberreraiketaren metodoan oinarrituta, euskarak igaro dituen garapen-garaiei buruz datu ugari jaso ditzakegu, hizkuntzaren oraina ere ulertzeko baliagarri direnak.

Mitxelenak erein zuen lur zabalean bildutako uzta esanguratsua da bai kopuruz bai kalitatez, maisu handiak ezarri zituen oinarrien sendotasunaren seinale. Hizkuntzaren alor guztietan, fonologiatik lexikora, aurreratu dugu gure ezagutzan, eta aspaldi iritsi zitzaigun garaia ikertutakoa zabaltzeko, gurean ez ezik baita mundu osoan ere. Euskarak beti jakin-min itzela piztu baitu atzerrian, dela bere gramatikaren ezaugarriengatik, dela Europako hizkuntzen mapan eta haren historiaurrean duen toki bereziarengatik. Ingelesez ez dira gutxi orain arte argitaratu diren lanak, Mitxelenaren garaitik hasita, baina haien artean gero eta gehiago dira erreferentziazko lantzat hartzeko modukoak, tartean gramatika deskribatzaileak edo euskararen historiari buruzko monografiak, guztiak gaur egun nazioartean oso erabiliak, hala nola José Ignacio Hualde eta Jon Ortiz de Urbinaren *A Grammar of Basque* (New York, Mouton de Gruyter, 2003) edo Robert L. Trasken *The History of Basque* (Londres, Routledge, 1997).

Orain esku artean duzun liburua, irakurle, kate oraingoz labur horretako beste begi bat da, eta ez nolanahikoa. Izan ere, euskararen gramatika historikoa izateko orain arte egin diren ahalegin sakonenetako bat dela baiezta dezakegu gehiegikerietan erortzeko arriskurik gabe. Sintesitik duen bezala, ikerketa berri eta berritzailetik ere asko du, eta euskararen iraganera hurbiltzeko saio egiatan iradokitzaileak bildu ditu bere orrialdeetara. Liburua idatzi duen hizkuntzalari taldeak, bestalde, eskarmentua eta gaztetasuna batzen ditu: ikertzaile gazte askoren parte hartzea pozgarria da diziplinaren jarraikortasunaren ikuspegitik.

Euskararen historiak ez du inoiz izan horrenbeste aditu, eta gaurko joerari eusten bazaio, aurrerantzean ere hizkuntzaren iraganari buruzko ikerketak luzaz egingo dira. Asko dago oraindik argitzeko, ezagutzen dugun historia luzea baita, baina are luzeagoa ezagutzen ez duguna, historiaurrearen laino artean galduz joan zaiguna. Behin iraganean murgilduta, garbi izan behar dugu dena ezingo dugula berreskuratu (aurkakoa pentsatzea frustraziorako bide ezinago zuzena da), baina era berean egia da egungo gure lanabesek eta ereduek euskararen bilakabidea nolakoa izan ote zen irudikatzeko eszenatoki arrazoizkoak eskaintzen dizkigutela, beste hizkuntzetan gertatutakoa aintzat harturik eta *tipologia diakronikoa* deritzonak ematen dizkigun datuetan oinarrituta.

Oraindik neurri batean zein bestean argitzeke dauzkagun hainbat esparrutan Mikel Martínez-Areta UPV/EHUko irakasleak koordinatu duen lan honek euskalaritzaren tradiziorik onenean bikain errotutako ideia eta proposamen berri ugari dakarzkigu, euskararen garai zaharretan izandako edo izan zitezkeen aldaketak hobeto ulertu eta azaltzen lagun diezaguketenak. Ez da meritu makala, gaian sakontzeko euskarak ez baitigu maizenik zuzeneko lekukotasun gehiegi ematen, eta horregatik iturririk ezak zurtz sentiarazten du askotan ikertzailea. Hizkuntzalaritzaren egungo garapen-mailak garbi ezartzen ditu, baita gurean ere, ongi arrazoitutako hipotesi sendoen eta hala-holako azalpenen arteko mugak. Euskararen historiari (eta batez ere historiaurreari) dagokionez, bigarrenek merezi zutenaren gainetiko oihartzuna lortu izan dute sarriegi. Orain aurkezten duguna bezalako lan akademiko serioek, ordea, euskararen iraganaren ezagutzan aurrera irmoki egiteko aukera ematen dute. Norabide horri, hain zuzen, eutsi behar diogu, euskararen gramatika historiko osoa mamitu eta gauzatuko badugu. Hara iristeko oraindik bidean goazela, ziur nago liburu honek (*Basque and Proto-Basque. Language-Internal and Typological Approaches to Linguistic Reconstruction*) helburutik oso gertu utziko gaituela.

Foreword

Blanca Urgell
(Minister of Culture, Basque Government)

In recent decades, considerable progress has been made in research on the Basque language and its history. Today, we understand the development of Basque better than ever before, and Bascologists have proposed solid and fruitful hypotheses. Research methods and tools have progressively improved and been standardized in order to obtain the richest possible information from the language's past. Luis Michelena stated —contradicting on this point Antoine Meillet, whom he otherwise considered his master— that Basque is in no way a language without history, and that its genetic isolation —i.e. its lack of known relatives other than Aquitanian— does not exhaust the paths of explanation for historical change. Quite the contrary: building upon both loan words borrowed from other languages and the method of internal reconstruction, we can collect plentiful data about the different periods of development undergone by the Basque language, which can also be useful to understand the present of the language.

The harvest gathered in throughout the extensive land sown by Michelena is considerable in both quantitative and qualitative terms, which proves the solidity of the bases established by the great master. From phonology to lexicon, we have increased our knowledge in every realm of the language, and the time has long since come to divulge what has been discovered, not only among Bascologists but on a wider stage. In fact, Basque has always triggered curiosity beyond our borders, be it due to its grammatical features, or to its special position within the map of the languages of Europe and its prehistory. Since Michelena's times, the amount of works published about Basque in English is considerable, among them a growing number of reference books including internationally well-known works such as *A Grammar of Basque* (New York, Mouton de Gruyter, 2003), edited by José Ignacio Hualde and Jon Ortiz de Urbina, or Robert L. Trask's *The History of Basque* (London, Routledge, 1997).

The book that the reader has in his/her hands is another link in this still relatively short chain, and not just any link. In fact, we can safely assert that it represents one of the most serious attempts made thus far at a historical grammar of Basque. Whilst it has an element of synthesis, it also brings together a great deal of new and innovative research, and thus constitutes a really suggestive attempt to bring us closer to the past of the language. The group of linguists who have written the book combines experience and youth, and indeed, the participation in

it of so many young scholars is very good news for the continuation of the discipline. The history of Basque has never had as many experts as it has today, and, if this trend continues, research on the past of the language will certainly be carried out for a long time in the future. There are many questions still to be resolved, as the known history of the language is long. However, the history that we do not know, that which is lost in the mists of time, is even longer. Once we plunge into the past of the language, we must bear in mind that we shall never recover everything (thinking otherwise is the shortest route to frustration). But it is also true that the linguistic tools and reconstruction models that we have at our disposal today allow us the possibility to conceive, within certain limits, reasonable scenarios as to how the development of the language may have occurred, taking as a model what has been done in the parallel research of other languages, and based on the data provided by the so-called *diachronic typology*.

This work, coordinated by the Assistant Professor of the University of the Basque Country (UPV/EHU) Mikel Martínez-Areta, brings us new ideas and proposals, well entrenched in the best tradition of Bascology, on a number of unresolved questions, thus helping us to better understand and explain the changes which took place or may have taken place in the old periods of the language. This is no mean feat, as most of the time the Basque language offers little direct evidence, and that lack of sources often causes the scholar to feel orphaned. The current state of linguistics clearly establishes, also within our field, the limits between well-argued strong hypotheses and unsound explanations. Worryingly, however, with regards to the history, and especially the prehistory, of Basque, the latter have too often achieved undeserved prominence. By contrast, serious academic works, such as the book we are now presenting, allow us to make firm progress in our knowledge of the past of the Basque language, the right direction to be moving in if we aspire to the broader project of a whole historical grammar of Basque. While we are still on the way, I am quite sure that this book (*Basque and Proto-Basque. Language-Internal and Typological Approaches to Linguistic Reconstruction*) will bring us close to that aim.

Table of Contents

Notes on the Contributors ..... *xv*
Abbreviations ..... *xvii*
Authors, texts and other written sources ..... *xvii*
Basque Dialects (according to Bonaparte's 4th classification) ..... *xxii*
Linguistic terms ..... *xxiii*
Frequently cited works ..... *xxv*
Frequently employed journals ..... *xxv*
Introduction (M. Martínez-Areta) 1
1. The Basque Language today. Achievements and challenges (A. Barreña, A. Ortega, E. Amorrortu) 11
1.1. Introduction 11
1.2. Sociolinguistic data of the Basque Country (*Euskal Herria*) 11
1.3. Factors affecting the increase in the knowledge of Basque 16
1.4. Positive attitudes toward Basque 19
1.5. The situation today: fresh challenges in the light of achievements 23
1.5.1. Language use and language transmission 23
1.5.2. Fostering affective links with Basque 28
1.5.3. Welcoming immigrants to Basque 28
1.5.4. The presence and development of Basque in the new technologies 29
1.6. Conclusion 29
2. Basque Dialects (M. Martínez-Areta) 31
2.1. Introduction 31
2.2. The Basque dialects today and in the historical past 31
2.2.1. How different are Basque dialects? 31
2.2.2. What are the Basque dialects? 34
2.2.2.1. Biscayan 35
2.2.2.2. Guipuscoan 43
2.2.2.3. Alavese 45
2.2.2.4. Northern High Navarrese 46
2.2.2.5. Southern High Navarrese 49
2.2.2.6. Labourdin 52
2.2.2.7. Western Low Navarrese 55
2.2.2.8. Eastern Low Navarrese 57
2.2.2.9. Souletin 59
2.2.2.10. Some other features: A-B-B-B-A and A-A-A/B-B-B alignments 63
2.2.3. How many Basque dialects are there? 66
2.3. The origin of Basque dialects 70
2.3.1. The secular view: tribes of Antiquity, Church dioceses, historical dialects 70
2.3.2. Lacombe (1952 [1924]), Uhlenbeck (1947 [1942]). Biscayan vs. the other dialects 73
2.3.3. Michelena's (1987 [1981a]) *éuscaro* or Common Basque 73
2.3.4. Zuazo's (2010a) innovating foci 75
2.3.5. Camino (2011), Lakarra (2011f) 78
2.4. Common Basque on the map 81
2.5. Conclusion 86
3. External History. Sources for historical research (K. Ulibarri) 89
3.1. Introduction 89

3.2. The Roman period. Aquitanian or archaic Basque and linguistic contact 92
3.2.1. Epigraphic evidence 93
3.2.2. Latin loan words 95
3.3. Medieval Basque. Centuries of lights and shadows 96
3.4. The Modern Era. Texts in Basque 98
3.4.1. 15th century – 1600. Archaic Basque 99
3.4.2. 1600 – 1745. Old Classical Basque 102
3.4.3. 1745 – 1887. Early Modern Basque 105
3.4.4. 1887 – 1968. Late Modern Basque 110
3.4.5. 1968 – Today. The unified Basque language (*euskara batua*) 112
3.5. Basque dialectology 113
3.6. Toponymy 114
3.7. Conclusion 116
4. Phonetics and Phonology (A. Egurtzegi) 119
4.1. Introduction 119
4.2. The Basque phonemic inventory: now, then, and before then 120
4.2.1. Modern Basque dialects 120
4.2.2. Michelenian stages 122
4.2.3. Lakarran Old Proto-Basque 124
4.3. Vowels 125
4.3.1. Main vowels 125
4.3.2. Nasalized Vowels 126
4.3.3. Distribution of the Souletin /y/ 127
4.3.4. Vocalic processes 130
4.3.4.1. Alternations between /i/ and /u/, /e/ and /o/ 130
4.3.4.2. Different vowel risings 131
4.3.4.3. Vowel lowerings 133
4.3.4.4. Vowel addition 133
4.3.4.5. Vowel deletion 135
4.3.4.6. Vocalic reciprocal metathesis 137
4.3.5. Diphthongs 137
4.3.5.1. Falling diphthongs 138
4.3.5.2. Rising diphthongs 139
4.4. Consonantism 140
4.4.1. Sonorants 140
4.4.1.1. Rhotics 140
4.4.1.2. Laterals 142
4.4.1.3. Nasals 144
4.4.1.3.1. /n/ 144
4.4.1.3.2. The development of /m/ 145
4.4.2. Stops 146
4.4.2.1. Voiced stop series 147
4.4.2.2. Voiceless stop series 148
4.4.2.3. Aspirated voiceless stop series 150
4.4.3. Fricatives 151
4.4.3.1. Different kinds of aspiration 151
4.4.3.1.1. Voiceless glottal fricative /h/ 151
4.4.3.1.2. Nasalized /h̃/ 152
4.4.3.2. The emergence of the labio-dental fricative /f/ 153

4.4.3.3. Sibilants 154
4.4.3.3.1. Voiceless fricative and affricate sibilants 154
4.4.3.3.2. Voiced alveolar sibilants 155
4.4.4. Palatal series 156
4.4.4.1. Palatal segments and their usage 156
4.4.4.2. Assimilations 156
4.4.4.3. Outcomes of **j-* 157
4.4.4.4. Depalatalization 158
4.4.5. Some consonantal processes 158
4.4.5.1. Assimilation 158
4.4.5.2. Dissimilation 160
4.4.5.3. Metatheses: reciprocal metathesis, perceptual metathesis and repercussion 160
4.5. Phonotactics 162
4.5.1. Word and syllable initial positions 162
4.5.2. Final position and coda tautosyllabic clusters 163
4.5.3. Heterosyllabic clusters and medial position 163
4.6. Accentuation 166
4.6.1. Central-western accentual type 167
4.6.2. Northern Biscayan accentual type 167
4.6.3. Souletin accentual type 168
4.6.4. Old Accentuation 169
4.7. Conclusion 171
5. Root Structure and the Reconstruction of Proto-Basque (J. A. Lakarra) 173
5.1. Introduction 173
5.2. The standard reconstruction of Proto-Basque 174
5.3. Reasons for change 177
5.4. A new theory for the reconstruction of Proto-Basque. The monosyllabic root 181
5.5. Theory of the root and formal etymology in the Proto-Basque lexicon 184
5.6. From the canonical form to diachronic holistic typology 189
5.7. The canonical form and the reconstruction of the Basque verb 192
5.8. Canonical form, word families, and the reconstruction of the Old PB phonology 196
5.9. Changes in the canonical form 202
5.10. Some conclusions and remaining tasks 204
5.11. Appendices 208
5.11.1. Tables 208
5.11.2. Etymological appendix 211
6. Noun Morphology (E. Santazilia) 223
6.1. Introduction 223
6.2. Description of today's Basque noun declension system 223
6.3. The article, definiteness and number in noun declension 227
6.3.1. The grammatical cases 228
6.3.2. The local cases 235
6.4. The origin of the primary cases 239
6.4.1. The grammatical cases 240
6.4.1.1. The ergative 240
6.4.1.2. The genitive 242
6.4.1.3. The dative 245
6.4.1.4. The instrumental 246

6.4.2. The local cases 247
6.4.2.1. The inessive 248
6.4.2.2. The allative 257
6.4.2.3. The ablative (and prosecutive) 261
6.5. Secondary cases 266
6.5.1. Cases built upon the allative 267
6.5.1.1. The directional allative 267
6.5.1.2. The terminative allative 268
6.5.2. Cases built upon the genitive 269
6.5.2.1. The sociative 269
6.5.2.2. The motivative 272
6.5.2.3. The destinative 272
6.6. The pseudo-flectional morphemes (or non-cases) 273
6.6.1. The local genitive 273
6.6.2. The essive/translative 276
6.6.3. The partitive 277
6.7. Conclusion 280
7. Demonstratives and Personal Pronouns (M. Martínez-Areta) 283
7.1. Introduction 283
7.2. Demonstratives 283
7.2.1. The paradigms of the standard language 283
7.2.2. Differences across dialects and historical developments 285
7.2.3. Peculiarities of case forms as compared to the noun declension 288
7.2.4. The demonstrative → article phenomena 294
7.2.5. The internal structure of deictic roots and the origin of the tripartite deixis 296
7.2.6. The category of number on demonstratives 299
7.3. Personal pronouns 301
7.3.1. 1st and 2nd person pronouns 301
7.3.2. Peculiarities of the case forms of 1st and 2nd p. weak pronouns 302
7.3.3. Weak vs. strong forms 306
7.3.4. The Linschmann-Aresti Law 312
7.3.5. 3rd person pronouns. Pronominal, anaphoric, and other uses 316
7.4. Conclusion 321
8. Non-Finite Verbal Morphology (M. Padilla-Moyano) 323
8.1. Introduction 323
8.2. Verbal periphrases in Basque 324
8.3. Tense and Aspect suffixes of non-finite verb forms: origins and variation 326
8.3.1. About the verbal radical (*-Ø*) and participle 326
8.3.1.1. A relative chronology of non-finite verbal forms 326
8.3.1.2. Classification of Basque verbs 327
8.3.1.2.1. The *-i* class 328
8.3.1.2.2. The *-n* class 328
8.3.1.2.3. The *-tu* class 329
8.3.1.2.4. The rest of the verbs 329
8.3.1.3. Expansion of the participle in western Basque 333
8.3.1.4. Reinforcement of the radical in eastern Basque 334
8.3.2. The third non-finite verbal form: the gerund 336
8.3.2.1. On the Basque gerund 336
8.3.2.2. The gerund, suffix by suffix 339

8.3.2.2.1. *-te* 339
8.3.2.2.2. *-tze* 339
8.3.2.2.3. *-eta* 340
8.3.2.2.4. *-keta* 340
8.3.2.2.5. The complex suffix *-(t)zaite* 341
8.3.3. The prospective participle 342
8.4. The Basque periphrastic verb: History and Pre-History 345
8.4.1. Combining non-finite forms with auxiliaries: TAM categories 345
8.4.2. Addition of further elements in periphrastic conjugation 348
8.4.2.1. Remoteness in present perfect and past 348
8.4.2.2. Marking the verbs as a sentence focus 348
8.4.2.3. Expressing progressive aspect: *ari*, *ibili*, *egon*, *jardun* 349
8.4.2.4. Further auxiliaries 350
8.4.3. Historical changes in the TAM system 351
8.4.4. Origins of the Basque periphrastic verb 352
8.4.5. The influence of Romance models 354
8.5. Conclusion 356
9. Finite Verbal Morphology (B. Ariztimuño) 359
9.1. Introduction 359
9.2. Background and brief description 359
9.2.1. Previous works 359
9.2.2. General traits 360
9.3. Person and number agreement 363
9.3.1. Person 365
9.3.2. Number 368
9.4. Tense-Aspect-Mood 373
9.4.1. Tense and Aspect 374
9.4.1.1. Specific morphemes 374
9.4.1.1.1. Present 374
9.4.1.1.2. Past 375
9.4.1.1.3. Hypothetical 376
9.4.1.2. Morpheme order and ergative displacement 377
9.4.1.3. Inherent Aspect: past, present and future 377
9.4.2. Mood 378
9.5. Valency and diathesis 381
9.5.1. *Nor* 383
9.5.2. *Nor-Nori* 385
9.5.3. *Nor-Nork* 387
9.5.4. *Nor-Nori-Nork* 389
9.5.4.1. Dative insertion 392
9.5.4.1.1. Dative flag(s) 392
9.5.4.1.2. Alternation of the present tense vowel 393
9.5.4.2. Causativization as dative insertion 394
9.5.5. More on the origin of the diathesis in the Basque verb: was Basque a passive language? 394
9.6. Allocutive agreement 397
9.6.1. The use of allocutive forms 399
9.6.1.1. Morphosyntactic triggering and constraints 399
9.6.1.2. Dialectal extension 400

9.6.2. The *hiketa* 401
9.6.2.1. Morphological changes 402
9.6.2.1.1. Addition of the addressee agreement 402
9.6.2.1.2. Stem suppletion 403
9.6.2.2. Phonological changes 405
9.6.2.2.1. Initial *d-* and the *-i-* insertion 405
9.6.2.2.2. Tense vowel and the *-i-* insertion 407
9.6.3. The *zuketa* and *xuketa* 408
9.7. Auxiliary verbs and synthetic verbs 410
9.7.1. Differences 410
9.7.1.1. Paradigmatic differences 410
9.7.1.2. Differences in meaning 412
9.7.1.3. Differences in morphology 414
9.7.2. Auxiliaries from synthetics, synthetics from Serial Verb Constructions (SVC) 416
9.8. Final remarks 417
9.8.1. Dialectal variants and reconstruction 417
9.8.2. Diathesis and agreement 422
9.8.3. SVCs and a possible origin of the Basque synthetic verbs 424
10. Word order (U. Reguero-Ugarte) 429
10.1. Introduction 429
10.2. OV vs. VO. Sentence constituent order in Basque 430
10.2.1. Syntactically conditioned constituent order 430
10.2.1.1. Affirmative sentences 430
10.2.1.2. Negative sentences 434
10.2.1.3. Sentences with a heavy O 436
10.2.1.4. Subordinate clauses 439
10.2.2. Pragmatically conditioned constituent order 440
10.2.3. Theoretical views on the sentence constituent order in Basque 443
10.3. Word order within the Noun Phrase 445
10.3.1. NPo vs. PrN 446
10.3.2. DetN vs. NDet (Det = Article, Demonstrative, Numeral, Interrogative) 447
10.3.3. GN vs. NG 449
10.3.4. RelN vs. NRel 451
10.3.5. AdjN vs. NAdj 453
10.4. Hypotheses about the word order in Proto-Basque 454
10.4.1. Trask (1977, *HB*) 455
10.4.2. Gómez (1994) and Gómez & Sainz (1995) 456
10.4.3. Lakarra (2005, 2006b) 457
10.4.4. Martínez-Areta (2011, 2012) 458
10.5. Conclusion 460
References..... 461
General Index 487

Notes on the Contributors

Esti Amorrortu is Associate Professor in the Department of Modern Languages and Basque Studies at the University of Deusto. Her research interests include Basque sociolinguistics, with a special focus on language attitudes towards Basque. Email: esti.amorrortu@deusto.es

Borja Ariztimuño is a PhD student in the *Linguistics* program of the University of the Basque Country (UPV/EHU), and Research Assistant and Instructor in the Department of Basque Studies and General Linguistics of the same university. His research focuses on the diachronic study of the Basque verbal morphology, etymology, and general typology. Email: borja.ariztimuno@ehu.es

Andoni Barreña is Associate Professor in the Department of Spanish Language at the University of Salamanca. His research interests include bilingual language acquisition and minority languages. Email: andoni@usal.es

Ander Egurtzegi is a PhD student in the *Linguistics* program of the University of the Basque Country (UPV/EHU), and Research Assistant in the Department of Basque Studies and General Linguistics of the same university. His research focuses on the history and prehistory of Basque phonology, as well as on the description and analysis of phonetic processes. Email: ander.egurtzegi@ehu.es

Joseba A. Lakarra is Professor of Basque Philology at the University of the Basque Country (UPV/EHU), editor of the journal *ASJU*, and member of the Royal Academy of the Basque Language (*Euskaltzaindia*). His research includes text edition, philological analysis of Basque texts, historical linguistics, etymology and general typology, but has particularly focused on the reconstruction of Proto-Basque taking the analysis of the canonical root as the starting point. Email: joseba.lakarra@ehu.es

Mikel Martínez-Areta is Assistant Professor of Basque Philology at the University of the Basque Country (UPV/EHU) and co-editor of the linguistic journal *Mikroglottika*. His research focuses on the historical morphology and syntax of Basque, and on general typology. Email: josemiguel.martinez@ehu.es

Ane Ortega is Associate Professor in the Department of Language Learning at *Begoñako Andra Mari* Teacher Training College. Her research interests include the didactics of language in bilingual educational programs and language attitudes. Email: aortega@bam.edu.es

Manuel Padilla-Moyano is a PhD student in the *Basque Linguistics and Philology* program of the University of the Basque Country (UPV/EHU), and Research Assistant in the Department of Basque Studies and General Linguistics of the same university, as well as a member of the research laboratory IKER UMR 5478 (CNRS). His research focuses on the edition and philological analysis of Eastern Basque texts, diachronic dialectology of Eastern Basque, and historical sociolinguistics. Email: manuel.padilla@iker.cnrs.fr

Urtzi Reguero-Ugarte is a PhD student in the *Basque Linguistics and Philology* program of the University of the Basque Country (UPV/EHU), and Research Assistant in the Department of Basque Studies and General Linguistics of the same university. His research focuses on Medieval Basque and the origin of dialectalization, as well as on the Navarrese dialects of the archaic and old periods. Email: urtzi.reguero@gmail.com

Ekaitz Santazilia is a PhD student in the *Linguistics* program of the University of the Basque Country (UPV/EHU), and Senior Lecturer in the Philology and Language Learning Department of the Public University of Navarre (UPNA/NUP). His research focuses on Basque philology, Basque historical morphology and general typology. Email: ekaitz.santazilia@unavarra.es

Koldo Ulibarri is a PhD student in the *Basque Linguistics and Philology* program of the University of the Basque Country (UPV/EHU), and Research Assistant in the Department of Basque Studies and General Linguistics of the same university. His research focuses on the edition and philological analysis of Basque texts, as well as on the diachronic dialectology of Western Basque. Email: koldo.ulibarri@gmail.com

Abbreviations

Authors, texts and other written sources

Ag. = Juan Bautista Aguirre (1742 – 1823). Guipuscoan priest and writer born in Asteasu. He composed several religious works which display an unusual command of the language. His most relevant text is *Confesioco eta Comunioco Sacramentuen gañean Eracusaldiac* (1803).

Amill. = J. B. Amilleta (1639 – 1700). He wrote a doctrine in the variety of Antzuola (ca. 1678), a Guipuscoan town located in the Biscayan-speaking area.

Añ. = Pedro Antonio de Añibarro (1748 – 1830). Biscayan Franciscan born in Villaro-Areatza. He spent the last 40 years of his life in Zarautz (Guipuscoa). Añibarro wrote many works, among them *Gueroco Guero* (a translation into Biscayan of Axular's *Guero*). Along with Moguel, he is considered to be the father of Biscayan literature.

Arg. = Pierre Argaignarats (17th c.). Priest of Ziburu (Labourd) who wrote two works in Basque. Only one of them has been preserved: *Devoten Breviarioa* (1665).

Ax. = Pedro de Aguerre y Azpilicueta "Axular" (1556 – 1644). *Axular* is the name of the house where he was born, in Urdazubi, to the north of the Baztan Valley (Navarre, but in the Labourdin-speaking area). He was the parish priest of Sara (Labourd) from 1600 until his death. He only wrote the *Guero* ('After'), printed in 1643, an ascetic work which soon became a model of linguistic elegance. Even today, Basque is sometimes referred to as "Axular's language".

Azk. = Resurrección María de Azkue (1864 – 1951). He was born in Lekeitio (Biscay) and died in Bilbao. In 1919, he was appointed president of Euskaltzaindia, the Royal Academy of the Basque Language. Azkue cultivated narrative, poetry, drama and opera in Basque. In addition, he was a prominent Bascologist. His *Diccionario Vasco-Español-Francés* and his *Morfología Vasca* are two reference works in Basque lexicography and Basque grammar, respectively.

Barr. = Pedro Ignacio de Barrutia (1682 – 1759). Public notary born in Aramaio (north. Alava). He is the author of *Actto para la noche buena*, a short Basque drama whose language is peppered with striking archaisms.

Ber. = Juan de Beriayn (1566 – 1633). He was the Abbot of Uterga (Navarre), where he was born and died. He published two works: *Tratado de como se ha de oyr missa* (1621) and *Doctrina christiana* (1626), which are the oldest long texts in HN.

Bet. = Juan Perez de Betolaza († ca. 1600). Priest, native of Betolatza (north. Alava). He is the author of *Dotriña Christiana en romance y bascuence* (1596), the earliest printed text in Alavese (if it can be considered as such) and, along with *RS*, the first work printed in western Basque.

Bp. = Athanase de Belapeyre. Parish priest of Sohüta (Soule), his *Catechima laburra* (1696) is the first book actually written in Souletin dialect, since both Oihenart and Tartas introduced some linguistic features from other continental dialects. This work is partially translated from Bossuet's *Catéchisme des fêtes et autres solennités et observances de l'église*.

Cap. = Martin Ochoa de Capanaga (ca. 1600 – 1661). Biscayan priest from Mañaria. He translated from Spanish into Biscayan Ripalda's *Exposicion breve de la Doctrina Christiana* (1656).

Cb. = Agustín de Cardaberaz (Hernani, 1703 – Bologna 1770). Guipuscoan Jesuit and theologian, author of over fifteen works. Outstanding among them is *Eusqueraren Berri Onac* ('The good news of Basque'), printed in 1761, a method addressed to those who wish to "correctly write, read and speak in Basque".

CSM = *Cartulary of San Millán*, also known as the *Becerro Galicano*. A late-twelfth-century cartulary penned in the Monastery of San Millán (in the High Rioja), containing hundreds of Basque place and personal names, often preserved in archaizing forms, the linguistically most interesting of which date from the 10th, 11th and 12th centuries. A singularly important document contained in the cartulary is the *Reja de San Millán*, dating from 1025, in which over three hundred Alavese villages which paid annual dues in iron or cattle to the monastery are listed. The significance of the list for linguistic research is huge, as the place names display an abundance of archaic features such as aspiration.

Dch. = Bernart Dechepare (16th c.). Priest, native of somewhere near Donibane Garazi (Cize, Low Navarre). His *Linguae Vasconum Primitiae* (Bordeaux, 1545), a collection of poems of humanistic influence, is the first book printed in Basque. Dechepare's life must be viewed in the context of the political and religious intrigues which took place in the Kingdom of Navarre in the 16th century.

Duv. = Joan Piarres Duvoisin (Ainhoa, 1810 – Ziburu, 1891). Head of Customs by profession, he was a learned man, entrusted by Prince Louis-Loucien Bonaparte with the first translation of the whole Bible, and some other works. Among several achievements, he also wrote some literary works, began a Basque-French Dictionary, and translated Cb.'s *Aita San Ignacioren Egercicioen gañean afectoac* from Guipuscoan into Labourdin, under the title *Liburu Ederra* ('The Beautiful Book'), published in 1856.

D.Zeg. = *Doctrina Christiana* (1741), written in the variety of Zegama (Guipuscoa), possibly by Martin de Urteaga.

Eliç. = Francisco de Eliçalde (1646 – 1733), born in Gesalatz (Navarre). He is the author of *Apecendaco dotrina christiana uscaras* (1735).

E.Cib. = Joannes Etcheverry (17th c.), best known as Etcheverry de Ciboure. Classical Labourdin writer who published three books: *Manual devotionezcoa* (Bordeaux, 1627), *Noelac eta berce canta espiritual berriac* (ca. 1630) and *Eliçara erabiltceco liburua* (Bordeaux, 1636).

E.Sar. = Joannes Etcheverry (Sara, 1668 – Azpeitia, 1749), best known as Etcheverry de Sara. This classical Labourdin writer composed three works: *Lau-urdiri gomendiozco carta, edo guthuna* (1718), *Escuararen hatsapenac* (an apologetic work) and *Euscal-herriari eta Euscaldun guztiei escuarazco hatsapenac latin icasteco*, a method for learning Latin.

F.Bart. = Bartolomé de Madariaga (1768 – 1835). Biscayan Carmelite born in Markina, author of a three-volume sermonary and a treatise against dances.

Gam. = Juan Bautista Gamiz (Sabando, 1696 – Bologna, 1773). Born in Sabando (southeast. Alava), he became a Jesuit in Valladolid, and from 1728 to 1767 he lived in Pamplona-Iruñea, until King Charles III banned the Jesuits from Spain (that is why he died in Bologna, like Mendiburu). He wrote mainly in Spanish, but 238 lines written by him in Basque (two long poems and four short ones) have been conserved until our days.

Gar. = Esteban de Garibay y Zamalloa (Arrasate-Mondragón, 1533 – Madrid, 1599). Guipuscoan learned man, he came to be the royal historian and chronicler of Philip the 2nd. He did not write in Basque himself, but in his two works *Compendio Historial*... (1571) and *Ilustraciones genealógicas*... (1596), written in Spanish, he often reports valuable news about the Basque language. Especially important are the epic poems gathered by him from oral (hence archaizing) tradition, which typically refer back to the political conflicts and fights of the previous century in Biscay, Guipuscoa and Alava. He also compiled two series of proverbs.

Gç. = Bernard Gasteluçar (Ziburu, 1614 – Pau, 1701). Labourdin Jesuit and author of an ascetic book, *Eguia Catholicac, salvamendu eternalaren eguiteco necessario direnac* (Pau, 1686). He is considered to be one of the finest Basque poets of his period.

Haran. = Joannes Haraneder (ca. 1669 – ?), born in Donibane Lohizune (Labourd). He is the author of a manuscript in which the Gospels are translated into Labourdin, and of two further translations of well-known ascetic works, published in 1749 and 1750.

K&B. = *Kadet eta Bettiriño*. A manuscript written in West.-LN (from Hazparne, Labourd) to be dated ca. 1750, which narrates in dialogue Jesus' life and death. Its author could be Gratien Diharce d'Alçuyet.

Lar. = Manuel de Larramendi (Andoain, 1690 – Loiola, 1766). Guipuscoan Jesuit and writer who conceived a cultural project to make Basque a language of social and cultural prestige. His apologetic work (1728), his grammar (1729) and, above all, his *Diccionario Trilingüe del Castellano, Bascuence y Latin* (Donostia-San Sebastian, 1745) prompted the beginning of Guipuscoan literature, as well as the emergence of Modern Basque, symbolically situated in 1745.

Laz. = Juan Perez de Lazarraga (ca. 1548 – 1605). Man of letters from Larrea (northeast. Alava), he penned i.a. a bucolic novel and several poems in Basque, the manuscript of which was found —partially damaged and with several gaps— in 2004. His language is our most important source of knowledge about the Alavese dialect in ca. 1600.

Lcc. = Niccolò Landucci (16th c.). Italian humanist born in Lucca. His *Dictionarium linguae cantabricae* (1562, unpublished until 1958), the first dictionary of Basque (Spanish-Basque), is believed to reflect the language spoken in Vitoria-Gasteiz in the 16th century.

Lç. = Joannes Leiçarraga (1506 – 1601). Born in Beskoitze (Labourd, but linguistically East.-LN. territory), among other works he translated, upon the request of the Calvinist Queen of (the Low) Navarre, *Testamentu Berria* 'The New Testament' (1571) into a Basque which was aimed to be understood by all continental Basque speakers. Religious reasons prevented Leiçarraga's titanic attempt to build a (continental) Standard Basque from succeeding.

L.Elc. = Joaquin Liçarraga de Elcano (1748 – 1835). Navarrese priest born in Elkano, author of several religious books, the most relevant being *Urteco igande guztietaraco platicac* (1846), a compilation of sermons in South.-HN. Basque.

Mend. = Sebastián de Mendiburu (Oiartzun, 1708 – Bologna 1782). His birth town is located in northeastern Guipuscoa, but linguistically belongs to North.-HN. territory. He was a Jesuit who left a number of works, both published and handwritten. He came to be known as "the Basque Cicero".

Mic. = Rafael de Micoleta (1611 – ?). Biscayan priest born in Bilbao, author of *Modo breue de aprender la lengua vizcayna* (1653), a succint learning method of Biscayan Basque, also containing several dialogues and poems.

Mog. = Juan Antonio Moguel (Eibar, 1745 – Markina, 1804). Priest and writer, author of *Peru Abarca* (1782, published in 1880), which has long been considered the first Basque novel, although current literary theory prefers to interpret it as a pedagogical dialogue. His language became a model for eastern Biscayan writers.

Mst. = Martin Maister (18th c.). Parish priest of Ligi (High Soule). His work *Jesu-Christen Imitacionia* (Pau, 1757), a translation of Kempis' *Imitatio Christi* into Souletin, is considered the zenith of Classical Souletin.

Och.A. = Joseph Ochoa de Arin (17th – 18th centuries). His work *Doctrina Christianaren explicacioa* (1713), written in the variety of Ordizia, is one of the few records of Old Guipuscoan.

Oih. = Arnaud d'Oihenart (Maule, 1592 – Donapaleu, 1667). Souletin lawyer, historian and poet. He published i.a. two works in Basque in a single volume: *Les Proverbes basques, recuellis par le Sr. d'Oihenart, plus les poésies Basques du mesme Auteur* (Paris, 1657), i.e. a compilation of proverbs (reflecting an archaizing, eastern Basque) and several poems. In these, the language is not exclusively Souletin, but uses forms from other continental dialects.

Ol. = Bartolomé Olaechea (18th c.). Native of Gautegiz-Arteaga and chaplain in Laukariz —both in Biscay—, he is the author of *Doctrina cristianea* (Vitoria-Gasteiz, 1763).

Peñ. = Peñaflorida (Azkoitia, 1729 – Bergara, 1785). Francisco Javier de Munibe e Idíaquez, Count of Peñaflorida. Distinguished Guipuscoan man of letters, he created and fostered the *Real Sociedad Bascongada de los Amigos del País*, a cultural society inspired by the French Enlightenment. Among other works, he wrote the comic opera *El borracho burlado* (1764), with texts in Spanish and Guipuscoan Basque, and the Christmas song *Gavon-Sariac* (1762), entirely in Guipuscoan Basque.

Port. = Martin Portal. Winner of a literary contest in Pamplona-Iruñea, in 1610. The poem —80 lines long— with which he achieved this award reflects a language which is to be located somewhere near to the Middle Sakana or the Burunda (western Navarre), or perhaps even further south, beyond the Urbasa Mountain Range (Amescoas, Lana), where Basque faded away early.

Pouv. = Sylvain Pouvreau (Bourges, ? – Paris, ca. 1675). Of non-Basque origin, this French writer was a priest in Bidarte (Labourd) from 1640 to 1644. He composed i.a. a Basque-French dictionary, and translated several religious works into Basque.

RS. = *Refranes y Sentencias*. A collection of proverbs printed in Pamplona-Iruñea in 1596. Most of them reflect an archaizing language which is to be located in or near to Bilbao.

Tt. = Jean de Tartas (17th c.). Souletin writer, author of two printed ascetic books: *Onsa hilceco bidia* (1666) and *Arima penitentaren occupatione devotaq* (1672). He introduced into these some linguistic features of Low Navarrese and Labourdin, in order to get closer to non-Souletin continental speakers.

Ub. = Juan Antonio de Ubillos (1707 – 1789). A Franciscan born in Villabona (Guipuscoa), author i.a. of *Christau Dotriñ Berri-ekarlea* (1785), a version in Basque of Fleury's *Catéchisme historique*.

Urt. = Pierre d'Urthe (Donibane Lohizune, 1664 – London, ca. 1725). Calvinist Labourdin writer. For religious reasons, he had to go into exile in England, where he wrote three works published after his death: *Biblia Saindua. Testament Çaharra eta Berria*, *Grammaire Cantabrique basque* (ca. 1712) and *Dictionarium Latino-Cantabricum* (ca. 1715).

VJ. = *Viva Jesus*. An anonymous Christian doctrine written and printed at some point of the first half of the 17th century (perhaps ca. 1640) somewhere in southwest. Biscay.

Volt. = Voltoire (17th c.). Man of letters who wrote *L'interprect ou traduction du François, Espagnol & Basque* (Lyon, ca. 1620), a learning method with dialogues, grammatical paradigms, proverbs and a small dictionary. The dialect reflected in it is Labourdin.

Zav. = Fray Juan Mateo de Zavala (Bilbao, 1777 – Zarautz, 1840). Biscayan Franciscan and writer, author of several fables and of *El verbo regular vascongado del dialecto vizcaíno*, an essay on the Biscayan verb.

Basque Dialects (according to Bonaparte's 4th classification)

Aesc = Aescoan. Subdialect of West.-LN spoken in the Aescoa Valley, in northern Navarre, to the west of Salazar. In spite of Bonaparte's classification, Aescoan exhibits some features which can be considered as peninsular.

Al = Alavese. Peninsular dialect formerly spoken in Alava, included within B by Bonaparte. In the 16th century, it was spoken in the northern strip of this territory, even up to Vitoria-Gasteiz and perhaps partially further south. Nowadays, it is restricted to the towns of Legutio and Aramaio. Traditionally, Alavese varieties have been regarded as a branch of Biscayan, but recent research tends to consider it an independent (historical) dialect with a relevant irradiation focus in Vitoria-Gasteiz.

B = Biscayan. The dialect spoken in Biscay (except in the western third of the province, where Basque is not spoken), and in the western strip of the province of Guipuscoa, mostly corresponding to the Deba Valley. It is highly idiosyncratic, and has several features which are not shared by any other Basque dialect.

East.-LN = Eastern Low Navarrese. The dialect spoken in the east of the Low Navarre, although its area also spreads to the northwestern part of this territory, and even enters northeastern Labourd. In the south, it also penetrates into the Salazar Valley in the High Navarre. It is Dechepare's dialect, and some of its linguistic features define it as an eastern dialect. In the northeastern corner of Low Navarre, the varieties of the Country of Mixe (Basque *Amiküze*) have a marked identity.

G = Guipuscoan. The dialect spoken in most of Guipuscoa (except at the northeastern end, where local varieties are linked to Navarrese, and in the Biscayan-speaking Deba Valley), as well as in some neighboring valleys of High Navarre. Due to its central location —it is surrounded on all sides by Basque-speaking areas—, Guipuscoan is an innovating, vigorous dialect.

L = Labourdin. The central dialect spoken in western Labourd (and in the Baztan Valley in the High Navarre, according to most classifications). In fact, Western and Eastern Low Navarrese are also spoken in the east of Labourd. In spite of its current limited geographical and demographic extent, Labourdin has historically been considered to be a very elegant dialect, and has consequently been fostered in much of the literary tradition.

North.-HN = Northern High Navarrese. The dialect spoken in the north and west of High Navarre (except in the Burunda, neighboring Guipuscoa), and at the northeastern extreme of Guipuscoa. North.-HN can be regarded as a central dialect.

R = Roncalese. A subdialect of Souletin (in Bonaparte's classification) or an independent dialect (in Azkue's classification) which was spoken in the Valley of Roncal (northeastern corner of Navarre) and which comprehended the varieties of Uztarroze, Urzainki and Bidankoze. It was already languishing in the 19th century, and its last speaker (from Uztarroze) died in 1992. It contained plenty of idiosyncratic features, of high interest for linguistic research.

S = Souletin. The dialect spoken in Soule (and in the Roncal Valley according to Bonaparte's classification). Its tendency towards archaism, together with its high degree of inner unity, makes Souletin a well-defined dialect with a strongly marked identity. Due to its rich tradition, Souletin is the most relevant dialect of eastern Basque, in contrast with its small number of current speakers (ca. five thousand).

Sal = Salazarese. Subdialect of East.-LN spoken in the Salazar Valley, in the High Navarre (to the west of the Roncal Valley). In spite of Bonaparte's classification, some of its linguistic features can be considered as peninsular. Today, it is quasi-extinct.

South.-HN = Southern High Navarrese. The dialect historically spoken in the central part of the High Navarre (hence the southern area of the Basque-speaking area), surrounding Pamplona-Iruñea on all sides. As a dialect, it was alive and somewhat vigorous at the time of Bonaparte's visits and research, but since then, Spanish has replaced it.

West.-LN = Western Low Navarrese. The dialect spoken in the west of the Low Navarre, except in the north (where East.-LN is spoken). It also comprehends a region of southeastern Labourd, the town of Luzaide-Valcarlos (in High Navarrese territory) and the Aescoa Valley of the High Navarre.

CB = Common Basque. The language from which all dialects would have branched, and from which the linguistic diversity of the different dialects would have originated. Michelena (1987 [1981a]) dated its existence to the 5th-6th centuries AD. It is important to bear in mind, however, that there are a number of important innovations which have occurred after that period but which have reached all dialects.

PB = Proto-Basque. The language which can be to some degree reconstructed by means of linguistic arguments (mainly internal reconstruction). It is prior to CB. Although there is absolutely no trace for its dating, it is sometimes estimated to have existed in the last centuries of the pre-Christian Era.

SB = (Modern) Standard Basque. The variety created and refined since the Congress of Arantzazu (1968) to enable understanding between the Basque speakers of different regions.

Linguistic terms

A Agent
Abl Ablative
Abs Absolutive
Adj Adjective
Adv Adverb, Adverbial
Affirm Affirmative
AL Allocutive
ALF Allocutive Familiar
All Allative
Alloc Allocutive
ALP Allocutive Polite
Aor Aorist
Aquit Aquitanian
Arc Archaic
Art Article
AUX Auxiliary (Verb)
AV Auxiliary Verb
A.CP Agentive Complement
Bearn Bearnese
Cast Castilian

Caus	Causative
Cent	Central
Comm	Common
Comp	Complement
COND	Conditional
COP	Copula
CPL	Complementizer
CS	Causal
c.	Century
Dat	Dative
Def	Definite
Det	Determiner
DF	Dative Flag
Dial	Dialectal
Dim	Diminutive
Dst	Destinative
East	Eastern
EP	Epenthesis
Erg	Ergative
F	Feminine
Fr	French
FUT	Future
G	Genitive
Gasc	Gascon
Gen	(Possessive) Genitive
GER	Gerund
HAB	Habitual
HYP	Hypothetical
IMP	Imperative
Indef	Indefinite
Ines	Inessive
Instr	Instrumental
INTR	Intransitive
IPFV	Imperfective
IRR	Irrealis
Lat	Latin
Lcv	Locative
L.Gen	Local Genitive
M	Masculine
Mot	Motivative
MV	Main Verb
N	Noun
NF	Non-Finite
NOMNL	Nominalizer
North	Northern
NP	Noun Phrase
NPRS	Non-Present
NPST	Non-Past
O	Object
Occ	Occitan
P	Patient (Agreement)
p	Person
PFV	Perfective
Pl	Plural
Po	Postposition
POT	Potential
PPSP	Personal Pronoun Suffixing Process
Pr	Preposition
Pre	Prefix
PROG	Progressive
Pron	Pronoun
Pros	Prosecutive
PROSP	Prospective
Prox	Proximate
PRS	Present
PST	Past
PTC	Particle
PTCP	Participle
Ptv	Partitive
p.n.	Proper Noun
R	Recipient
RAD	Radical
Rel	Relative (or Relative Clause)
Resp	Respect
RLV	Relativizer
Rom	Romance
S	Subject (or Subject Intransitive)
SB	Subordinator
SBJV	Subjuntive
Sg	Singular
Soc	Sociative
South	Southern
Sp	Spanish
Suff	Suffix
SVC	Serial Verb Construction
T	Topic
TAM	Tense-Aspect-Mood
Term	Terminative
TMP	Temporal
TR	Transitive
V	Verb
V_F	Finite Verb
V_{NF}	Non-Finite Verb
West	Western

Frequently cited works

ContTAV	*Contribución al Estudio y Edición de Textos Arcaicos Vascos* (= Sarasola 1990 [1981])
EDH	*Euskal Dialektologiaren Hastapenak* (= Txillardegi & Aurrekoetxea (coord.) 1983)
EMH	*Euskal Morfologiaren Historia* (= Azkarate & Altuna 2001)
FHV	*Fonética Histórica Vasca* (= Michelena 1977 [1961])
GB	*A Grammar of Basque* (= Hualde & Ortiz de Urbina (eds.) 2003)
HB	*The History of Basque* (= Trask 1997)
MV	*Morfología Vasca* (= Azkue 1969 [1923-25])
OEH	*Orotariko Euskal Hiztegia / General Dictionary of Basque* (= Euskaltzaindia 1987-2005)
TAV	*Textos Arcaicos Vascos* (= Michelena 1990 [1964a])

Frequently employed journals

ASJU	*Anuario del Seminario (de Filología Vasca) Julio de Urquijo*
BAP	*Boletín de la Real Sociedad Vascongada de los Amigos del País*
FLV	*Fontes Linguae Vasconum*
RIEV	*Revista Internacional de Estudios Vascos*

Introduction

Mikel Martínez-Areta
(University of the Basque Country, UPV/EHU)

Much of the literature on the history of the Basque language —the only surviving language of non-Indo-European origin in western Europe— has dealt with its origin and possible relatives. A list of the many attempts to relate it to another language or language family would be exasperatingly long, and most probably not exhaustive. Leaving aside the obvious relationship of historical Basque to Aquitanian, Trask (1995a) refuted all of them, or more precisely, denied that any of the alleged relationships had been proven so far, and in Trask (*HB*: 358-429) he made a magnificent *reductio ad absurdum* of the methods which had been used in such attempts, arguing that such a methodology can lead us to prove whatever linguistic connection we wish (as he demonstrated with Basque and Hungarian). Since then, such attempts have not ceased, some of them unprecedented. The Basque language, however, continues to be a language isolate, and there is no indication that it will cease to remain as such in the near future.

Unfortunately, this ceaseless effort of historical research to establish unprovable theories has often diverted attention from paths which might have proven successful for making progress in the elucidation of its prehistorical aspects. As a result, Basque diachronic linguistics suffers from a historical delay in comparison with the equivalent discipline in many other languages or language families. In particular, progress made in diachronic cross-linguistic typology since the seventies has not been systematically applied to Basque until very recently, and then only in literature mostly written in Basque, Spanish or French. The purpose of this book is precisely to fill this gap, to provide an account of this situation, and to turn the spotlight on this possibility for research.

The starting point for this purpose, then, will be that Basque has as yet no proven relative, and hence no possibility for linguistic comparison beyond its geographical boundaries (for the Basque-Iberian question, see below). Our sources of knowledge are limited, and any honest attempt to explore the past of the language ought to start from this assumption. Written texts do not appear until roughly 1500, and then only in particular dialects and geographical areas. The attestations from the five centuries prior to this, mainly consisting of onomastic material in documents written and transmitted in medieval monasteries, do not provide any crucial data which can be of help for historical grammar. The same holds true for the Aquitanian inscriptions engraved on stone in the Roman

period (ca. 1st, 2nd and 3rd centuries AD), and for the place names provided by Greek and Latin authors from Antiquity.

With such a scarcity of sources, Bascologists have had to maximize the utility of the information provided by the linguistic contact with Latin and Romance. It cannot be a coincidence that the first scholar who tried to carry out systematic historical linguistic work on Basque in the modern sense of the word, Hugo Schuchardt (1842 – 1927), was a Romanist. Since Schuchardt's work,[1] Basque linguistic studies have made much progress, but, as I have said, it is hard to escape the impression that serious diachronic work has been delayed by misconceptions, ucronic conceptions and clichés, most of them derived from its status as a language isolate and from the fact that much research has focused on the search for relatives. Thus, it was not until the 1950s that the structuralist method was first employed to establish some basic game rules in historical reconstruction. This enabled Basque diachronic linguistics to catch up with the functionalist-structuralist methodology which flourished in the post-war period, but, indeed, in the diachronic research of other language families such as Romance, Germanic, Slavic or Indo-European linguistics, the Neo-Grammarians and their phonetic law had already provided a useful tool for reconstructing analysis as early as the end of the 19th century.[2]

These functionalist-structuralist approaches entered the Basque discipline thanks to André Martinet (1908 – 1999, who was not a Bascologist himself, but was the spearhead of the functionalist school), René Lafon (1899 – 1974), and especially Luis Michelena (1915 – 1987). Basque linguistics in the second half of the 20th century, even after Michelena's death and in fact until the present day, is dominated by this prominent figure. Before I move on to comment on his work in some detail, however, I would like to observe that, in my opinion, Basque historical grammar has still much to profit, in several respects, from a deeper reading of Lafon's work. Some crucial keys to understanding the formation of both the nominal and the verbal inflections, such as the lack of grammaticalized categories of number and diathesis, were already pointed out by him, were later ignored —or very superficially considered— for decades, and only in recent years are they being timidly recovered. The explaining potential of these views in the light of the progress made in recent typological research is, as I see it, enormous.

Luis Michelena's work, and particularly his *Fonética Histórica Vasca* (1961), is the sieve through which any reconstructing hypothesis must pass be-

1 For a history of the linguistic work on Basque, cf. Trask (*HB*: 49-70).

2 It is perhaps worth recalling that Schuchardt vigorously opposed the Neo-Grammarian views.

fore being subjected to ulterior tests. It is with Michelena that diachronic research on Basque reaches the maturity necessary to locate any posed problem or hypothesis within the coordinates of the scientifically possible and reasonable. As Trask (*HB*: 67) pointed out, Michelena worked on practically every linguistic area: historical phonology, the verbal system, place names, personal names, etymology, syntax, genetic relations, Romance influences, dialectology, literary criticism, the word-accent, etc. And for practically all of these areas, his studies have been the starting point for further research in recent decades.

After Michelena's death, Basque diachronic linguistics came to an impasse which lasted for some time. Indeed, the impression somehow caught on that his work on historical issues was insuperable, and that little or no contribution could ever be made to the field beyond his work.[3] In addition, the 70s and 80s brought the generative fever, and —perhaps somewhat unexpectedly in a discipline like Basque linguistics— scholars' interest drifted towards synchronic issues.

In spite of all that, Basque philologists continued their meticulous labor, which brought major lexicographic and editing achievements. The accentual diversity of Basque, a realm in which Michelena insisted that research was still in its infancy, was carefully studied by Hualde in the nineties, from both a synchronic and diachronic point of view. And since 1995, upon the solid basis of Michelena's reconstructing paradigm, Lakarra has developed another paradigm which attempts to go even further back in the history of the language, taking the study of the canonical root as the keystone. This diachronic enterprise was also strongly favored by foreign Bascologists such as Roger Larry Trask (1944 – 2004), Rudolf de Rijk (1937 – 2003) and Georges Rebuschi. The premature death of the first two was particularly unfortunate for the field, as one cannot avoid the impression that they left us just when a discussion largely generated and fostered by them was beginning to become interesting, and that their participation in it would have been of the utmost relevance.

As a consequence of that perseverance, Bascologists are more optimistic today than several years ago about the possibility of obtaining fruitful results in diachronic research, and in the last ten years there has been a renewed interest in the field among scholars and graduate students (the contributors in this volume are a good token of this). In comparison to giants like Lafon or Michelena —to use the well-known metaphor attributed to Bernard of Chartres—, we may be dwarfs, but we are dwarfs standing on their shoulders. In addition to having their work already completed, we have more sources and more typological information than they ever had access to. The situation in this respect is also more

3 With regard to historical phonology, Trask (*HB*: 6) stated: "[s]ince Michelena, little remains beyond tidying up the details".

favorable than that of 16 or 18 years ago, when Trask's *The History of Basque* (Trask 1997) and *Towards a History of the Basque Language* (Hualde & Lakarra & Trask (eds.) 1995) were respectively published. Within the realm of Basque philology and linguistics, several texts have been discovered, Lazarraga's manuscript being the most important. Dialectal research has filled some gaps that we had about archaizing dialects, and a number of inscriptions from Antiquity containing Basque onomastic data have also been found since Gorrochategui's canonical work on the issue (Gorrochategui 1984). While all this has not drastically changed the overall picture and we still have to face the same, at times demoralizing, scarcity of sources, the sum of all of these elements has enlarged our knowledge of several important aspects.

As Romance and Indo-European diachronic linguistics show, apart from the number of records we have, a prerequisite for progress is often simply that enough time elapses for scholars to process and arrange the information at their disposal. A Bascologist of 2012 is lucky to have, after many years of laborious work, a *General Dictionary of Basque* (Euskaltzaindia 1987-2005), in 16 volumes, with copious records from the written tradition for each lexical item, and Yrizar's *Morfología del Verbo Auxiliar Vasco* (Yrizar 1991-2008), in 15 volumes, which is virtually an encyclopedia of the Basque auxiliary verb. Similarly, in addition to the discovery of the new texts mentioned above, other texts that we already had but which had either remained unedited, or only very unsatisfactorily so, have now been properly edited and more profoundly discussed.

In addition to the progress made within Basque philology and linguistics, cross-linguistic typological research at an international level has also considerably advanced in the last four decades. Since the 1970s, new linguistic universals have been set up, some that were already known have been better attested, some others refuted, and still others have given rise to endless controversies between opposing views. By the same token, frequent grammaticalization paths have been discovered and/or discussed. Trask (1977, 1979, i.a.) was a pioneer in the use of these resources for diachronic research on Basque, but the bibliographical flood has been particularly copious in the last two decades. As a result, today we have essential handbooks such as Corbett (2000), Siewierska (2004), Bybee & Perkins & Pagliuca (1994), Heine & Kuteva (2002), etc., as well as many more descriptive monographs on particular languages or language families. A path of research which is particularly promising in relation to Basque, and which has begun to be systematically exploited by Lakarra and some of his disciples, is the study of serial verb constructions and their grammaticalization, especially into different postpositions, which in turn may become suffixes. Here again, without the assistance of contributions like Lord (1993) and Aikhenvald & Dixon (eds., 2006), such an enterprise would prove much more difficult.

By contrast, a field which has not rendered any helpful results for Basque diachronic research, at least as far as grammar is concerned, is its comparison with the Iberian language. The theory that Basque and Iberian are originally the same, or at least relatives, lost ground in the course of the 20th century, first with the decipherment of the Iberian script by Gómez-Moreno in the second quarter of the century, and then with the work, critical of their relationship, by Tovar and Michelena in the 1950s. Some serious Iberists such as Orduña (2005) and Ferrer i Jané (2009) —and any reader who is acquainted with the literature which is often published in this realm will understand the relevance of the adjectivization of 'Iberists' here— have recently revived that line of research.[4] However, their analysis has mainly focused on numerals and some lexical items. Hence, even if we accepted their analysis, and/or even if we believed that Basque and Iberian were to some degree relatives, it would still continue to be true that Iberian has not yet been of any help in the elucidation of any aspect related to the historical grammar of Basque.[5] Consequently, Iberian data are barely mentioned throughout the entire work.

This book is an attempt to present a general picture, for both Bascologists and typologists —especially those interested in diachronic phenomena—, of the

4 Against the possibility of such a comparison, see Lakarra (2010b). Orduña explained the similarities between Basque and Iberian numerals by assuming that the former borrowed them from the latter, whereas for Ferrer i Jané those numerals would be cognates descending from a common language. However, Orduña (2011) has recently adhered to Ferrer i Jané's view on this particular point.

5 The Iberian attestations which have been related to Basque and which could theoretically provide some help for grammatical comparison are very few, and extremely dubious. Thus, an *ekiar* found in Caminreal (Teruel) and an *ekien* found in Mendigorría (Navarre) have been related to western Basque *egian* '(s)he made'. More possible forms with this lexeme *eki-* 'make' are identified by Orduña (2010), who even suggests that an *eroki-*, allegedly a causative of *eki-*, might be the lexeme in two further inscriptions. Such a semantics certainly belongs to the narrow lexical scope which we expect to find in inscriptions, but the comparison with *egian* has, among others, the crucial problem that we reconstruct **(z)e-gin-a/e-n* for Basque, and hence we would expect the appearance of an intervocalic *-n-* —or at least *-h-*, which does not exist in Iberian— in *ekiar*, *ekien*, etc. Some other classical similarities with Iberian, such as the filiation suffix *-(t)ar* or the alternation *-n* / *-r* / *-Ø* are interesting and worth exploring, but still insufficient for helping us to better understand any grammatical aspect of Basque from a diachronic point of view. Another question which has not been settled is whether Iberian was a unitary language throughout the entire peninsular east —and if so to what degree—, or it only corresponded to a limited region within that area (perhaps that of the Contestani of Antiquity and some other tribes around them), and in other regions where Iberian inscriptions are found —such as those closest to the Basque-speaking territory— it was simply a *lingua franca*, as proposed by de Hoz (1993).

current situation of that research. Its structure, then, resembles —or at least is aimed to resemble— that of a historical grammar. Extra-linguistic issues such as population genetics or paleo-anthropologic research are left aside. While smatterings of Basque grammar, philology and culture can be useful for grasping some details, the work is intended to be comprehensible for as broad a public as possible. Hence, each chapter explains the linguistic features and facts in historical Basque, and the ample section of Abbreviations at the beginning contains a guide to the dialects and the main linguistic sources that appear throughout the work.

Chapter 1 (by Barreña & Ortega & Amorrortu) is a description of the Basque language today, in which an accurate account of the number of speakers and their percentages in each administrative region, and of the evolution of those rates during the last decades, is given. The socio-political background underlying all those data is likewise discussed, as well as the historical factors which have led to them. Finally, the sociolinguistic aspects of today's linguistic situation are described. Speakers of Basque are classified into different types, and an attempt is made to account for their subjective attitudes towards the language.

In Chapter 2 (by Martínez-Areta), the degree of intelligibility between Basque dialects is explained, they are enumerated one by one and the basic information about each one (its geographical situation, approximate number of speakers, degree of idiosyncrasy, the antiquity of the oldest sources, etc.) is provided. The issue of how many dialects there are is also briefly touched upon, but the diachronically most relevant section discusses at what chronological point a Common Basque can be postulated as the origin of all dialects, and how the branching process of those dialects may have occurred. Before entering the linguistic discussion proper, in Chapter 3 (by Ulibarri) a short external history of the Basque language is given. Since this also serves the purpose of describing the majority of the sources available to us for the historical research of Basque, this external history is complemented by two sections about dialectology and toponymy, which are two additional sources of linguistic evidence.

Chapters 4 and 5 deal with phonetics, phonology and root structure. In Chapter 4 (by Egurtzegi), the phonemic inventory of Basque in the historical period is described, and then the reconstructions proposed by Michelena for Proto-Basque, and by Lakarra for Pre-Proto-Basque, are set out. First, this is done for the vocalism, then for the consonantism. Finally, phonotactics and the old accent systems are also discussed. Chapter 5 (by Lakarra) is not only an explanation of the root structure in diachronic terms, but also of how its study has served as a starting point for the development of a new reconstructing paradigm at all levels. It also contains two appendices. The first consists of three tables in which all attested root structures are analyzed with respect to all their possible

theoretical combinations. The second is a sample of 50 proposed etymologies which exemplify most of the root structure processes defended throughout the chapter.

We move next to the morphological part of the volume, which comprehends chapters 6 to 9. Chapter 6 (by Santazilia) is a study of the Basque noun morphology, which sets out the noun paradigm of historical Basque and the structure of its NP, and then goes on to list the explanations proposed so far for every aspect concerning the formation of that paradigm: first the number/definiteness axis, and then the individual cases, classifying these into primary cases (grammatical and local), secondary cases (those built upon the allative and the genitive), and pseudo-flectional morphemes or non-cases. In Chapter 7 (by Martínez-Areta), demonstratives —in both their determinative and pronominal functions— and personal pronouns are discussed. The discussion about the origin of the former is considered in connection with the origin of the noun paradigm. As for the latter, of special relevance is the alternation between weak and strong (or even hyperstrong) forms, the syntactic/pragmatic conditioning of that alternation, and, in the case of the genitive forms, the origin, functioning period and decline of the Linschmann-Aresti Law.

Chapters 8 and 9 deal with the verbal morphology, splitting it into the study of the non-finite verb and the finite verb, as is usual in analysis, according to the structure of the Basque verb. In Chapter 8 (by Padilla-Moyano), first the morphological formations —by means of an ample array of allomorphs of diverse origin— of the participle, of the gerund (both built, at least originally, upon the verbal root), and of the prospective participle are explained. Secondly, the author discusses the origin of the different kinds of periphrases which resulted from the possible combinations of all those non-finite verbal forms with the different kinds of auxiliaries —built upon verbal lexical roots which over the course of time became auxiliary—, thus giving rise to a tense/aspect/mood system whose diachronic development from Common Basque to contemporary Basque is extremely interesting. In Chapter 9 (by Ariztimuño), in turn, the morphological structure and origin of the finite verbal forms —i.e. of both the synthetic verbs and auxiliary verbs of periphrastic constructions— are described and discussed. For that purpose, the different categories which intertwine in that structure, such as person, number, tense, aspect, mood and valency are dissected and analyzed both individually and in relation to the rest of the parameters. Then, one of the most characteristic features of the Basque finite verb, allocutive agreement, is covered, and finally the differences between auxiliary and synthetic verbs are discussed, both synchronically and diachronically. The chapter ends with a set of interesting conclusions and indications for future research.

In Chapter 10 (by Reguero-Ugarte), which closes the volume, the main features of Basque word order are summarized, first those corresponding to clause-level phenomena, which can be syntactically or pragmatically conditioned, and then those within the phrase. Some of the latter show diachronic and/or diatopic variation in the attested data, and hence a directionality with respect to the OV vs. VO dichotomy is worth analyzing, as it can provide us with important clues about the changes which have occurred in all the other parameters, including the phonological and morphological ones. Finally, the main hypotheses about the word order in Proto-Basque, which inevitably have to resort to internal reconstruction —focusing especially on the morphology of the finite verb—, are put forward.

For place names which have an equivalent in English, the form in English has been used. Obviously, it is not always easy to draw a line between those which have an English form and those which do not. Clear cases are e.g. Navarre and Biscay, but what about Cizain or Aescoan? In general, I have interpreted that the names of the provinces (*herrialde* in modern Standard Basque) and dialects can be given in English, and here this work differs from Trask (*HB*), in which e.g. the Souletin dialect is called *Zuberoan*. By contrast, town and village names are given in their official Basque forms (except in some cases such as Bilbao, where the Basque official form *Bilbo* sounds somewhat awkward). Elements from intermediate levels, such as subdialects, varieties or valleys are given in Basque (*Sakana*, *Deba Valley*, etc.), unless we have considered that they have an English equivalent. Obviously, this decision may be arbitrary in some cases (as is the case of *Roncalese*, *Salazarese*, *Aescoan*, *Cizain*, etc.). In the cases of town names in which the Basque and the Spanish/French official names are completely different (e.g. Basq. *Iruñea* / Sp. *Pamplona*, capital of Navarre) —not mere spelling (e.g. Basq. *Azkoitia* / Sp. *Azcoitia*) or pronunciation variants (e.g. Basq. *Sara* / Fr. *Sare*), or variants which are somewhat similar to each other due to their common origin (e.g. Basq. *Beskoitze* / Fr. *Briscous*)—, both variants are usually given.

As for the cited authors, the names of those corresponding to researchers mentioned in the references are spelled as they themselves signed their work. So are in general those corresponding to the linguistic sources, but this is not always done, and a justification for the election is not systematically provided. To give but one example, Lazarraga will be spelled with <z>, and its abbreviation will be Laz., even though he signed as *Laçarraga*, in order to strengthen the distinction from Lç. (corresponding to Leiçarraga), as they will both be frequently cited throughout some chapters of the text.

By contrast, the extracts of texts taken from the written tradition are given, for practical reasons, in the spelling system of Standard Basque, following the

procedure used in the *General Dictionary of Basque*. However, data from medieval documents, or even from later documents but which do not belong to the literary corpus proper —such as those found in archives, diplomatic collections, etc.—, are given in their original spelling.

Some of the abbreviations for dialects, authors and works have been inspired by the system used in Trask (*HB*) or in the *General Dictionary of Basque* (Euskaltzaindia 1987-2005), but the resulting whole is original, and is hence detailed in the Abbreviations. My final resolution comes across a couple of possible duplicities. One is B, which by itself stands for Biscayan, but in PB stands for Proto-Basque, in CB for Common Basque, and in SB for Standard Basque; the other is S, which stands for Souletin when referring to the dialect, but for Subject when referring to the clause constituent, and for Subject Intransitive when referring to the argument type. Since the context makes it absolutely clear which one is involved in each case, I have not considered it necessary to take any measure beyond warning against them at this point.

The chronological terminology which is used throughout the work contains two basic terms. One is Common Basque (= CB), which is the diachronic projection of all Basque historical dialects, and hence the old common language prior to the dialectal fragmentation, from which all dialects branched. Michelena (1987 [1981a]) dated it to around the 5th-6th cc., and while there is no conclusive proof, it is commonly accepted since then that this might be close to the reality. This does not exclude, however, the existence of multiple innovations which have occurred after that period but which have reached all dialects. The other basic term is Proto-Basque (= PB), which would be a stage of the language prior to CB, commonly associated with Michelena's classical reconstruction of the consonant system. Much more arbitrarily, it is dated to the last centuries before the Christian Era, its chronological position being sometimes defined as the stage of the language immediately prior to the beginning of its contact with Latin. Throughout the volume, however, some authors stratify PB into different phases, which they can specify by adjectivizing PB with Old, Classical, Pre-, etc. In particular, Lakarra's theory of the root is to be located in a phase previous to PB, which can therefore be labeled as Pre-Proto-Basque or Old Proto-Basque.

This volume is a monograph published within the series of studies *Mikroglottika*, an international journal of minority language philologies printed by the Peter Lang Publishing Group, and co-edited by Raúl Sánchez-Prieto, Daniel Veith, and myself. As the editor of this particular volume, I would like to thank my co-editors of *Mikroglottika* Raúl Sánchez-Prieto and Daniel Veith for their encouragement in the production of this volume, and the publisher Peter Lang for their agreement on suggested ideas and their assistance in carrying them out.

The impression and a linguistic supervision of the text have been funded by the Department of Culture of the Government of the Basque Autonomous Community, to which the editor and the contributors of this volume wish to thank for their generous help. I am also thankful to Ruth Yates, for her patient reading and linguistic supervision of the text, and to Alphax Studio for the design of the figures and maps of Chapters 1 and 2. Likewise, I would like to mention that I belong to a number of research projects, within the framework of which this work has been carried out:

1) The research project "Monumenta Linguae Vasconum III-IV: historia, crítica y edición de textos vascos" [= FFI 2008-04516, FFI2012-37696], led by Joseba Lakarra and funded by the Ministry of Education and Science of the Spanish Government.
2) The research project "Historia de la Lengua Vasca y Lingüística Histórico-Comparada" (HLMV-LHC), Ref. GIC 10/83, IT 486-10, led by Joaquín Gorrochategui and funded by the Government of the Basque Autonomous Community.
3) The Training and Research Unit "Hizkuntzalaritza Teorikoa eta Diakronikoa: Gramatika Unibertsala, Hizkuntza Indoeuroparrak eta Euskara" (UFI 11/14), funded by the University of the Basque Country (UPV/EHU).

1. The Basque Language today. Achievements and challenges

Andoni Barreña (University of Salamanca)
Ane Ortega
('Begoñako Andra Mari' Teacher Training College, Bilbao)
Esti Amorrortu (University of Deusto)

1.1. Introduction*

The Basque language, called *euskara* by its speakers, is spoken by around 800,000 *euskaldunes* (speakers of *euskara*) who live on the north side (France) and south side (Spain) of the Western Pyrenees, in a territory called *Euskal Herria*, the country of the Basque language. *Euskal Herria* is divided into different political administrations: the Basque Autonomous Community or Euskadi (henceforth BAC), the Foral Community of Navarre, and the French Basque Country, which is part of the Department of the Pyrénées-Atlantiques.

The number of Basque language speakers has significantly risen over the past few decades, bearing in mind that only 30 years ago the Basque-speaking community was estimated to be of around 600,000 speakers.

This chapter describes the sociolinguistic development of the language over the last 30 years; examines the factors that have enabled the growth in the number of Basque speakers; reports on the language attitudes of the population towards Basque; and, finally, discusses the challenges the language faces in the light of what has been achieved so far.

1.2. Sociolinguistic data of the Basque Country (*Euskal Herria*)

The Basque Country as a whole has a total area of 20,947 km^2 with a population of just under three million inhabitants, according to official census figures (Aztiker and Soziolinguistika Klusterra 2011). Within this, the BAC has an area of 7,261 square kilometers and approximately 2100,000 inhabitants; Navarre,

* The research on attitudes towards Basque reported in this chapter was commissioned and funded by the Language Policy Office of the Basque Government (Amorrortu *et alii* 2009). Funding was also provided by the Spanish Ministry of Education and Universities (MICINN bid FFI2009-13956-C02-01, program FILO). The authors would like to thank these institutions, as well as all members of the public who participated in the study.

10,420 km^2 and roughly 600,000 inhabitants, and the French Basque Country, an area of 2,992 km^2 and about 250,000 inhabitants (cf. Map 1.1).

Map 1.1. Euskal Herria territory (the Basque Country) and population.

The extension of the Basque language has varied over time and, although the Basque Country today comprises the above territories, the area where Basque or its past forms are or were spoken is known to have been larger in the past, spreading east along the Pyrenees, as far as the high basin of the Garonne River. Other boundaries are controversial due to the scarcity of sources (see Núñez-Astrain 2003 for details).

The population of the Basque Country barely exceeded one million at the end of the 19th century, but the strong process of industrialization during that time, which in turn attracted a massive influx of Spanish immigrants into the BAC, led to a significant increase in population.

Massive Spanish immigration, together with the fact that Basque did not have official status until 1983, and the existence of long periods when the use of the language was prohibited, denigrated or rejected, caused the percentage of Basque speakers to drop to just 25% of the population by 1980.

Figures 1.1 and 1.2 show data collected by the official population censuses (1981, 1991, 2001, 2012) and the 'sociolinguistic surveys' (Language Indicator System of the Basque Country database) conducted by the Language Policy Office of the BAC Government (Basque Government 2003, 2008, 2009). These data reflect the evolution in the knowledge of Basque by the population of

the Basque Country over the last 20 years (Figure 1.1) —note that it is only for this period that global data for the whole territory are available— and linguistic competence in 2011 (Figure 1.2).

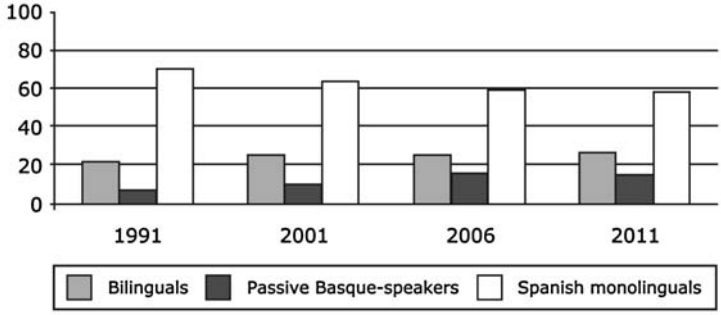


Figure 1.1. Evolution of knowledge of Basque in the Basque Country (1981-2011) (%): Basque speakers or bilinguals, passive bilinguals or bilinguals who understand but do not speak, and Spanish or French monolinguals. Source: Basque Government (2003, 2008, 2009, 2012).

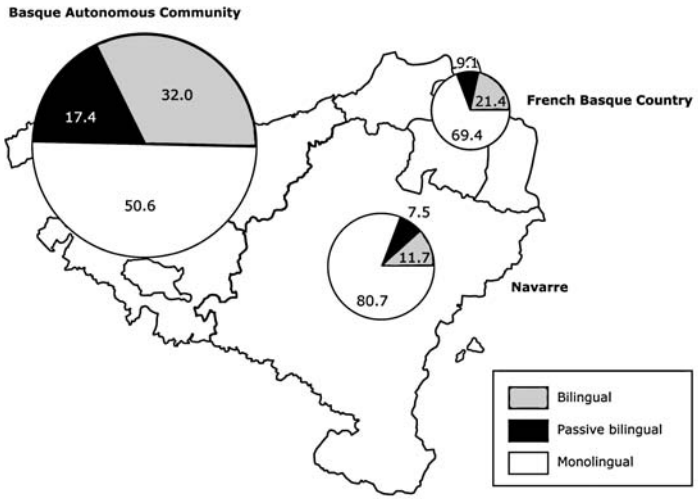


Figure 1.2. Linguistic competence in the three political administrations of the Basque Country (2011). Source: Basque Government (2012).

It is important to note that there is no monolingual Basque-speaking population, at least not one which is significant enough: there may be some elderly speakers with limited Spanish or French competence, but the percentage has no sociolinguistic relevance; on the other hand, there are also Basque monolinguals in the age group 0-4, but these will develop their second language throughout their childhood and, as a result, will cease to be monolingual at an early age. Thus, in terms of their language repertoires, Basque speakers are bilingual with Spanish or French.

The 5th Sociolinguistic Survey of the Basque Government (2012) shows the following data for the 15+ age group of the population of the Basque Country in 2011: more than a quarter (27%) is bilingual in Basque and Spanish or Basque and French and are, therefore, *euskaldunak* or Basque speakers; around one seventh of the population (14.7%) understands Basque but has difficulties speaking it, they are therefore passive speakers; and almost 6 out of 10 (58.4%) are Spanish or French monolinguals.

As can be observed from the data in Figure 1.1, between 1991 and 2011 the percentage of Basque speakers increases as follows: balanced bilinguals, from 22.3% to 27%, and passive bilinguals, from 7.7% to 14.7%.

However, this overall increase does not mean that Basque is developing in the same way in all three territories or political administrations. In the BAC, with data available from 1981, positive development is clear: balanced bilinguals have increased from 21.9% to 32% and passive bilinguals from 12.2% to 17.4%. In Navarre, the increase is much smaller: balanced bilinguals from 9.5% to 11.7% and passive bilinguals from 4.6% to 7.5%. And in the French Basque Country the opposite trend is observed, as the knowledge of Basque has diminished substantially: the percentage of balanced bilinguals has dropped from 33.1% to 21.4% and that of passive bilinguals has slightly increased, from 7.1% to 9.1%. Figure 1.3 shows the development of Basque speakers by territory. Note that the percentages for balanced bilinguals and passive bilinguals have been added together, from 1981 to 2011 in the BAC, and from 1992 to 2011 in the other territories.

The analysis of the data by age shows an important fact, that of the increase in knowledge of Basque, particularly in the younger generations. This increase is especially noteworthy in the BAC (Figure 1.4), where, for example, in the 16-24 age bracket, the knowledge of Basque rose from 34.1% in 1981 to 82.9% in 2011. During this time, knowledge of Basque also increases among the younger generations of Navarre, but much less radically, while it drops substantially in the French Basque Country.

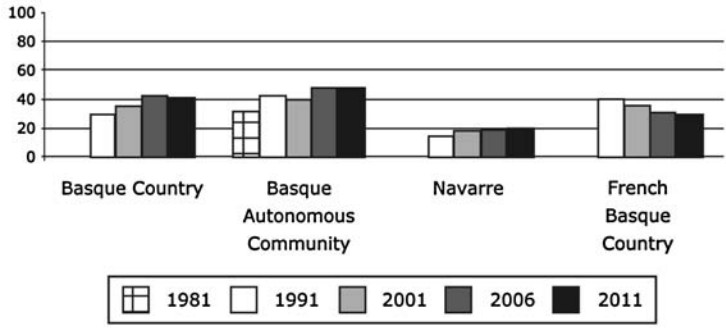


Figure 1.3. Evolution of knowledge of Basque in the different political administrations of the Basque Country (1981-2011) (%). Source: Basque Government (2003, 2008, 2009, 2012).

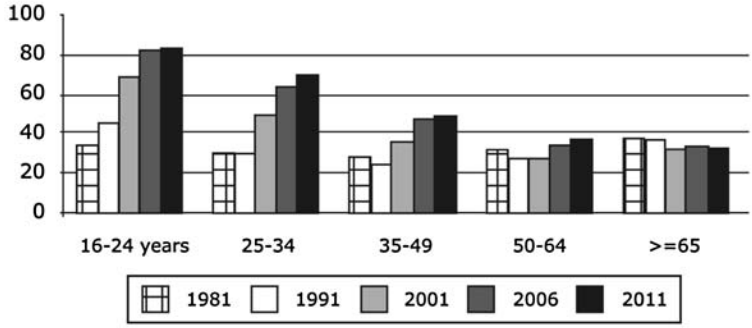


Figure 1.4. Evolution of knowledge of Basque in the BAC per age in population 15+: balanced and passive bilinguals (1981-2011) (%). Source: Basque Government (2003, 2008, 2009, 2012).

But language competence in Basque is not the only aspect that has improved in the last few decades. The minority language has also clearly gained domains of use since the beginning of the language policy which favors it, at least in the BAC. It is now used in the whole of the education system (from infant education to postgraduate university studies), in mass media (with two television stations, a daily newspaper and a great number of radio stations and magazines broadcasting exclusively in Basque), and in public administration.

However, use in informal settings and situations has not grown as expected (except for in Guipuscoa, the region with the highest percentage of bilinguals),

despite the big increase in the number of new speakers, and this is a cause for social concern, as will be addressed in §1.5. In fact, the percentage of Basque language use in colloquial interaction, measured in the streets all over the Basque Country, increased from 23.3% in 1989 to 32.7% in 2011 in Guipuscoa, but from only 3.9% in 1989 to 4% in 2011 in Alava, and in Biscay, from 8.1% in 1989 to 9.4% in 2011 (Soziolinguistika Klusterra 2012).

1.3. Factors affecting the increase in the knowledge of Basque

Among the factors that have influenced the increase in the number of new speakers of Basque and its better situation in terms of use, two important ones must be underlined. On the one hand, there was a powerful grass-root movement arising at the end of the Franco era, a movement of social and political contestation against the regime, which placed the loss of Basque and the rights of Basque speakers at the very center of its fight (Tejerina 1992): during the 1960s and throughout the 1980s, thousands of L1 speakers and non-speakers of Basque involved themselves in the promotion of Basque, creating Basque infant and primary schools, attending literacy and L2 Basque language evening classes, and engaging in all kinds of fund-raising and promotional activities whose main goal was the promotion of Basque language and culture in different settings. The beginning of democracy in Spain and the creation of the autonomous state allowed the establishment of a governmental language policy which implemented some of the measures already taken or promoted at grass-root level.

On the other hand, the introduction of Basque into the education system must also be highlighted as a crucial factor for the increase in the number of Basque speakers. Since Basque is a non-Indo-European language, and its linguistic distance from the dominant languages Spanish and French is large, the learning of the language in adulthood could not be expected to be enough for the recovery of the language. That is why the decision to make Basque mandatory in education has been so important and is probably the single most determining factor for the recovery of the language.

The Basque language, or *euskara*, had neither had any official status nor had it been a language of instruction until the 20th century. Even though there were some attempts to instrumentalize instruction in Basque and to make Basque the administrative language during the Spanish Second Republic (1931-1939), these efforts were abruptly aborted by the subsequent authoritarian administration. It is only from the sixties that the model of immersion in Basque was introduced through the *ikastolas* or Basque schools, which were initially illegal or semi-clandestine. As a consequence of the excellent academic results obtained by the

students coming from the first *ikastolas*, and the generalization of the immersion models in Basque Public Education from the democratic era in the eighties, both *euskara* and Basque language education underwent an important expansion and a growth in prestige.

In 1979, the Statute of Autonomy of the BAC was proclaimed, and in 1982, its namesake for Navarre (Law on the Reintegration and Enhancement of the Foral Regime of Navarre). The Law for *euskara* was passed in 1982 in the BAC and in 1986 in Navarre (Intxausti 1990). These laws favored and fostered the use of Basque both in administration and education and contributed to its prestige. However, the two autonomous communities —the BAC and Navarre— have followed different policies, due to qualitative differences in the regions and also to the level of commitment to Basque by their governments and the extent of their willingness to support its development.

In the case of the BAC, the administration has made possible, broadly speaking, linguistic policies which strengthen Basque. In education, there are three linguistic models to be chosen by parents: Model A in Spanish with Basque as L2; Model B with both Spanish and Basque as languages of instruction; and Model D in Basque. Note that Model D is in effect a mother-tongue model for Basque-speaking children and an immersion model for Spanish-speaking children.

The success of models D and B, both in general academic results and in language competence (in Basque, Spanish and English), has led to its generalization, as shown in Figure 1.5. The fact that it is the parents themselves who choose the bilingual models is indicative of their involvement in the process of revitalization and of the growing prestige acquired by the minority language.

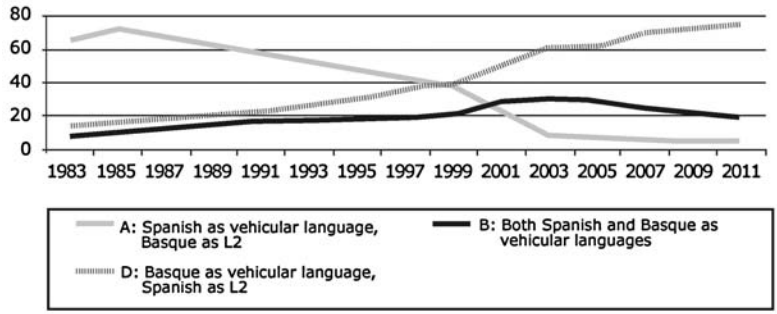

80
60
40
20
0
1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011
A: Spanish as vehicular language, Basque as L2
B: Both Spanish and Basque as vehicular languages
D: Basque as vehicular language, Spanish as L2

Figure 1.5. Evolution of enrolment at the age of 3 in the different linguistic models (A, B and D) in the BAC (1983-2011). Source: Berria (2011).

As can be observed in Figure 1.5, while the percentage of children enrolled in Model D has dramatically increased during the last three decades, Model A has clearly declined, even though the first language of children is still Spanish in most families in the BAC. Among the reasons that non-Basque speaking parents may have to enroll their children in immersion models, the following have been found (Amorrortu & Ortega 2009): commitment with the Basque revitalization movement and a wish for their children to integrate in the Basque-speaking community; the instrumental prestige gained by Basque, and the advantages these parents see in their children being bilingual; and the widespread belief that Basque is a very difficult language to learn as an adult, but easy for children, especially in an early immersion program.

It is the children of these bilingual models which account for the drastic increase in speakers discussed above. Whilst it is the parents who make the decisions in mainstream education, it should be noted that nowadays nearly 50% of university students choose to study in Basque.

The situation is different in Navarre. The Government of Navarre has adopted a zoning policy based on the different linguistic areas (Basque-speaking, mixed-speaking and non Basque-speaking), implementing different linguistic policies in each area. This has made possible, for example, the use of Basque in public education in the first two zones but not in the third one, a measure strongly contested by Basque language advocates in Navarre. On the basis of the above facts, the increase in the knowledge of Basque in Navarre has grown more slowly. In any event, the percentage of three-year-old children enrolled in model D had reached 35% in 2011 (Hik Hasi 2011).

In the French Basque Country, the Basque language has no official status or any form of governmental support, which places it in a really critical situation. Nevertheless, there is a grass-root movement which actively promotes the language and, with regard to education, there exist two types of Basque schools: the *ikastolas*, where Basque is the language of instruction and French is the L2, and Ikas Bi, a model in which both French and Basque are languages of instruction. Until recently, three-year-old enrolment in these schools was low, only 15.9% in 2006, but there has been a significant increase in more recent years, with figures reaching 30% in 2008.

In the territory where there has been clear progress, the BAC, the combination of individual and collective efforts, positive language policies towards the *normalization* of Basque, popular initiatives and support from people to Government policy, and the prestige achieved by the school system in the Basque language, are the cornerstones for the growth of Basque speakers and the increasing use of the language.

1.4. Positive attitudes toward Basque

The improvement in the general situation of Basque, at least in the BAC, is not only due to the commitment of society, but also to an improvement in the prestige of the language, as shown by sociolinguistic surveys and qualitative research on attitudes towards Basque.

The Sociolinguistic Surveys, which have been carried out every five years since 1981, clearly show that attitudes towards the promotion of Basque have improved, although in a different way across the three administrative territories. Surveys are useful indicators of general trends, but they have limited value for phenomena as complex as attitudes and the reasons for them. Qualitative studies, on the other hand, are still scarce but have yielded interesting results on how the different relevant groups perceive Basque and its recent developments. The following relevant profiles can be distinguished:

a) *euskaldun zaharrak* or 'old speakers' (traditionally, native speakers from Basque-speaking backgrounds). The number of L1 Basque speakers (including those who acquired both Basque and Spanish/French in early childhood) has increased from 13% in 1991 to 22% in 2006 in the whole of the Basque Country, according to the respective Sociolinguistic Surveys. They are the speakers who feel the closest attachment to the language and who have the most positive attitudes toward Basque and its promotion. Their past and future importance for the maintenance of the language is unquestionable.
b) *euskaldun berriak* or 'new speakers' (traditionally, Basque L2 speakers). It is estimated that more than 300,000 people have learned the language as an L2 in the BAC alone, where the total number of speakers has risen from 419,200 in 1991 to 755,640 in 2006 (Basque Government 2008), and mainly, but not only, thanks to the introduction of Basque in education through the immersion models, as discussed above. L2 speakers are especially strategic in the Basque case, given the low percentage of 'traditional' native speakers. This group will be further discussed in §1.5.
c) *erdaldunak*, or non-Basque speakers. Although the number of non-Basque speakers has decreased in the last two decades (from 65.9% in 1981 to 45.2% in 2006 in the BAC; Basque Government 2008), their number is still very high. This group, although apparently the one with the least attachment to Basque, has played an important role in its promotion, as shown, for instance, by the increasing number of children from Spanish-speaking families enrolled in immersion models (cf. Figure 1.5).

The makeup of attitudes towards Basque will be more precisely illustrated with the group of non-Basque speakers, still a majority of the population and

allegedly the one that could be expected to be emotionally the farthest from Basque. The attitudes of this group should be a good indicator of how Basque society regards *euskara* and Basque linguistic policy.

A recent study (Amorrortu *et alii* 2009) on the language attitudes towards Basque in the BAC used complementary methods to collect and analyze the data: qualitative data obtained in focus groups provided an in-depth understanding of a large array of attitudes towards Basque and its promotion, and then some of the most salient issues were measured quantitatively in a survey. The study confirmed that the attitudes toward Basque and its promotion are generally favorable. On the one hand, there is a virtually unanimous consensus on the idea that the language must be maintained, with statements such as 'Basque should not be left to die', 'I agree that Basque should be supported' and 'The use of the Basque language should be respected', and considerable support for concrete measures to achieve this, as shown by statements such as 'If a language in danger is to be saved, active measures for support must be taken, even if there is not absolute consensus about them', 'It is essential that public funds are used for the promotion of Basque' and 'The costs of the language policy in favor of Basque are reasonable' (Amorrortu *et alii* 2009: 280-300, 325-329). The quantitative study also showed three different groups of non-Basque speakers with regards to their positioning towards Basque language policy and their level of commitment to supporting it (Amorrortu *et alii* 2009: 235-304, Ortega & Amorrortu 2010): the *more-in-favor*, the *more-critical* and the in-between *yes-but* group, the latter being a very interesting group which will be further described later in this section. On the other hand, negative associations towards Basque that could be found forty years ago have been reduced drastically, among them the perception of Basque as a non-modern or primitive language (only 1.8% of the sample agreed totally with the statement 'Basque is a language of peasants' (Amorrortu *et alii* 2009: 269). This is thanks to the greater presence of Basque in public life both at the macro-level of institutional policy and media, and at the micro-level, in daily life.

As well as this, the study also reveals a great complexity of attitudes, stemming from different experiences, opinions, beliefs, and ideologies. As mentioned above, a very interesting result of the study has been the identification of what has been called the *yes-but* group; that is, those who despite being generally favorable towards Basque and its promotion, set limits towards what is reasonable and may question some aspects of Basque promotion. Most speakers in this profile are non-Basque speakers, live in mainly Spanish monolingual areas, and do not have access to the cultural world in Basque. Their limits to Basque-promoting policy are often expressed by means of complex, lukewarm and even

sometimes contradictory, ideas. Here are some of the issues they brought up and their ideas:

a) The yes-but position themselves in relation to Basque in particular ways. Their attitudes toward Basque and the language policy being carried out by the government are generally favorable, although they set limits, as shown in c). Regarding how they perceive the Basque language, while some consider it to be **the** language of the Basques, others believe Basque to be **one of the** languages of the Basques, together with Spanish or French:[1]

(1) *Well, I have always taken both Castilian and Basque to be the two national Basque languages, but on equal terms. (KA-B, 29)*

b) Most yes-but are aware that they do not have access to the cultural world in Basque and some wish they could access it:

(2) *I know that I am losing out because I cannot speak Basque. (BI-B, 13)*

c) They set limits to Basque language promotion and to governmental policy in general. Although they are generally in favor, they question measures they regard as extreme and set limits to some of the language policies implemented by the Basque government. Some of the ideas mentioned are: Basque should not be given priority over professional ability itself (3); the same language policy should not be implemented in all areas, and not the same should be expected from everybody (4); the recovery process is going too fast, there is a lot of pressure (5):

(3) *Well, I think... if I have to be operated on, I want the best doctor to do it, I don't give a monkey what language they speak. / I think it's what you were saying earlier, that when you go to the health centre you should have the option to be seen by a doctor who speaks Basque, that is important. But then, if we are talking about a surgeon... well, don't give me that, please.(BI-B, 121/123)*

(4) *I think it's fine that Basque is supported. It is culture, and all that is culture should be preserved and developed. But extending it to the whole of the population of Euskadi [the BAC], of Vitoria or Alava, which have a very different situation, or where what society wants is different, that is a very different thing. (GA-D, 70) [Note: Alava, and its capital city Vitoria, are the areas of the BAC with the lowest social presence of Basque].*

1 The translation from Spanish is ours.

(5) *In my opinion... the problem I see is that they have tried to reverse the effects of that time of repression too quickly and to go from a situation where hardly anybody spoke it, only those who spoke it at home, to making it a prerequisite for a number of jobs. It is perhaps this which has caused some rejection among people.(BA-C, 198)*

d) The yes-but are often people who have tried to learn Basque as an L2 at one point and, interestingly, feel that they need to justify why they did not succeed, often with regret.

(6) *I was born in 1965, it was the bad times [Note: reference to the Franco regime]. I did try to learn, but I was very bad at it, could not do it. But I have made sure that my children go to Basque schools. (ZA-D, 9)*

e) A number of simplifications and naïvetés that minimize the precarious situation of Basque are often present among the yes-but, for instance, considering that Basque can only have a happy future, in view of all the efforts taken to promote Basque so far, but failing to take into account the pressures that a language in the situation of Basque still faces. In addition, they often assume erroneously that knowing Basque guarantees its use. This misbelief will be addressed in more detail in §1.5.

(7) *Well, it's just a matter of our generation disappearing, isn't it? Because, once we disappear, our children will speak in Basque and will speak Basque to their children, they will learn at home, and everybody will speak Basque. (ZA-C, 122)*

The results of this study clearly show how important it is to get to know the attitudes of the relevant groups in order to take the right steps towards the normalization of language: policy makers need to know what society is willing to do, as well as how to foster positive attitudes.

More studies are needed on relevant groups, specifically on new speakers or *euskaldun berris*[2] and the native speakers or *euskaldun zaharras*. Another strategic group for the future of the language is, of course, the young.

2 At the moment, there is a recently started study on the attitudes of new speakers of Basque: "Euskaldunberrien jarrerak, motibazioak eta identitatea / Actitudes, motivaciones e identidad de los neohablantes del euskera" (2011-2014), funded by the Bizkailab initiative (Diputación Foral de Vizcaya and University of Deusto) (cf. Ortega *et alii* 2013).

1.5. The situation today: fresh challenges in the light of achievements

Following on from the description of the situation of Basque so far, this section attempts to discuss what has been achieved and what can be considered to be cause for concern, as well as the challenges Basque faces for the future. The following aspects will be examined: language use and the challenges for the different relevant groups, namely, the new speakers, the native speakers and the young; the importance of fostering positive attitudes towards Basque; the integration of immigrants and the presence of Basque in the new technologies.

1.5.1. Language use and language transmission

The dramatic increase in competence in Basque in the BAC, especially in the under-25 group as already discussed, is very good news for a language whose critical mass of speakers was so low at one point, that this in itself was a fundamental obstacle for Basque speakers wanting to use their language. However, as reported above, the major concern is that its use has not increased in general terms as much as could be expected. Whilst the use of Basque in the BAC, and in Navarre to a lesser extent, has understandably increased in domains traditionally banned to this language (education, the media, etc.), the concern relates to the small increase observed in the private domain, typically, in interactions among family and friends. Before discussing this further, a note of caution is due here against naïve expectations: it was never reasonable to expect an increase in use proportional to the increase in knowledge, as language use patterns are often slow to change. Moreover, the opportunities to speak are still not always present today. This is certainly the case in Spanish-speaking areas. In fact, it is not surprising that it is in Guipuscoa, the territory with the highest percentage of bilinguals, that Basque is most used now; the greater opportunities to speak it will no doubt have something to do with this.

Concerns regarding language use have mostly been flagged in relation to two groups: the younger generations and new speakers. With regards to the young, it is a widespread perception among adults that ‘the young speak Basque perfectly well but do not use it’, something that seems puzzling to many: after the great commitment and effort of society to support Basque education as the best way to extend the opportunities to learn the language, it seems astonishing that those who have had such an opportunity fail to use it. With regards to new speakers, it is an obvious challenge that those ‘gained’ by Basque make it an active language and find opportunities to use it. However, the authors of this pa-

per believe that the use of Basque should be looked at in a more global way and that the 'old' speaker profile and their language choices should also be examined, as well as what happens in the different sociolinguistic contexts and communicative situations. It should be taken into account that the contact situation with Spanish and French as a whole has changed a lot in a short time and new language-choice and language-use patterns are also being developed. We will next try to unpick some of these issues.

The younger generations are without any doubt fundamental for the present and future development of Basque, indeed much is expected of them. As said above, they are perceived not to use Basque as much as they could, or in other words, that they use Spanish quite extensively. This applies not just to those coming from Spanish-speaking families and environments, but also to those living in Basque-speaking areas. As stated before, their linguistic behavior is regarded with astonishment as well as dismay, because much is expected of these new Basque generations, who are already establishing patterns of language use and who are the ones that will transmit the language to the future generations. But is this perception true and fair?

In the first place, it is only fair to say that it is not really known how the young use their language(s) among themselves. Moreover, different groups in different areas may be displaying different language-use patterns and it should not be generalized that the youth perform as one group. Some studies (e.g. Esnaola 1999, Amonarriz 2008, Ibarra 2010) report that youth talk often features extensive code-switching, which may be perceived by many as using 'a lot of' Spanish and not the heritage language, and failing to see it as a code in itself, without mentioning the stigma that often accompanies mixed codes. While such dismay is understandable given how much there is at stake, one must bear in mind that the young of today are not the young of 30 years ago, just because their experience with the two languages has been so very different. There is a risk of prejudicing the behavior of the young by applying the values, hopes and fears of the older generation, very different to those of the new one. It is crucial, therefore, to find ways to collect reliable data. Bearing in mind that the in-group speech of young people is very sensitive to adult observation, data collected by techniques such as participant-observation should yield truly trustworthy results that may or may not confirm the general perception. And when the results are analyzed, it must be taken into account that youth speech is by definition 'transgressive' and that usage at the age of adolescence to early twenties may not be a solid predictor for language-use patterns in adulthood.

A very different kind of evidence may shed further light on the issue. The number of cultural products created by and for the young is actually quite staggering: the schools of *bertsolaritza* —traditional oral poetry based on improvisa-

tion— are full of children and young people; performing groups of Basque dance are numerous and extremely popular; music bands singing in Basque, from the most traditional to the most avant-garde and fusion are plentiful and are followed by thousands in their concerts. This evidence suggests that Basque does indeed have a very important presence among the young.

Is there not, then, cause for concern? There is cause… for research. The development of the young Basque speakers must be closely monitored. As well as getting to know how their uses develop, as advocated before, it is also crucial to discover how they feel about Basque and Spanish, what the images associated with their two languages are, what attitudes and beliefs they have, and what their motivations are for exercising their language choices. Regarding beliefs, for instance, there is already some evidence that the under-25 group, for whom the presence of Basque in schools and society has been 'normal', may be under the impression that Basque is no longer under threat and that their individual choices make no difference; what has been termed the *rosy future* belief (Amorrortu *et alii* 2009). Should this belief be confirmed, awareness-raising action must be taken urgently, because of its harmful effects for the fate of a small language like Basque in the globalized world. Another challenge that Basque faces in relation to the young involves the images that they associate with Basque. As the above-mentioned study shows, the language seems not to be sufficiently associated with youth culture and what is *cool*. Rock bands, for instance, are very important, precisely for bringing the heritage language closer to the world of experiences of the young, and should therefore be supported.

Another strategic group is the **new speakers of Basque**. The term is understood here as a wide-reaching notion which includes any Basque speaker who has acquired the language by any means other than family transmission. Therefore, this definition includes very different profiles in terms of age and mode of acquisition, most notably, two main ones: those who have learned Basque in youth or adulthood, and those who, coming from Spanish-speaking families and sociolinguistic environments, have attended immersion-in-Basque educational programs from the age of early socialization.[3] The reason why these speakers are not in principle excluded from the definition is because many of them show a relationship with the language which is similar to that of an L2 in some respects (such as attachment to the language, self-image as 'new speakers', linguistic features).

3 In the BAC virtually 100% of children are schooled by the age of 3. Note that the data for the 2009-10 school year show that, out of the total number of children starting in Infant Education, 71.13% did so in the Basque immersion model D, 23.89% in the bilingual model B, and only 4.98% in the Spanish model A (Basque Government 2009).

Why this group should be strategic is obvious, as the increase in the number of speakers in the last 40 years is largely due to them. For this reason, it is of the utmost urgency to understand their attitudes and what they are prepared to do in favor of Basque. Right now, however, very little is known about them. Answers to the following questions seem crucial:

- What are their motivations to speak Basque, what are the obstacles or difficulties they experience when communicating in Basque, and, what are the circumstances that encourage usage?
- Little is known about the identity configuration of new speakers and their type and degree of attachment to Basque. How important is Basque to their identity? Do they feel close to Basque, proud to speak the language and to be part of the Basque-speaking culture, do they feel they have gained something special by learning the language? What is its relation to their L1, Spanish or French?
- For many of these speakers being 'new' is regarded as somewhat lacking and they may not see themselves as rightful and legitimate speakers of the language. What are the factors affecting self-image and legitimation? Is it a question of the linguistic variable (many new speakers' variable is the standard *batua*, and not one of the vernacular dialects), a question of competence, the lack of opportunities to use the language on a regular basis...?

All of these ultimately lead to the most crucial question: will new speakers be active promoters of Basque, contributing to its maintenance and its development, and will they transmit Basque to their children (supposedly as well as Spanish/French)? Which profiles of new speakers will, and which ones will not? A recently launched three-year study (Amorrortu *et alii* 2011-2014, in progress, and the footnote in §1.4) intends to uncover the general picture and start answering some of these questions.

As previously argued, we believe that the issue of the use of Basque should be approached in a more global way, and should also include the **traditional native or 'old' speaker (*euskaldun zaharrak)***, understood loosely to be those who have learned the language at home. Native Basque-speaking communities have ceased to lose speakers in the BAC and, to a certain extent, in Navarre (a trend that did not begin until the change in the political situation in the late 70s). This is excellent news. And not only that, but the cohort of 'old' speakers has increased in the BAC, for two reasons:

a) The traditionally Basque-speaking group has consolidated due to generalized family transmission.
b) Many Basque-L2 parents have transmitted Basque to their children as their L1 or one of their L1s. This generation is mostly in the under-25 age

group and is, in fact, the first generation of standard/*batua*-L1 speakers, as it was the standard *batua* variety which many of these parents learned as an L2.

Despite the good news, 'old' speakers also face some challenges. In the BAC, 'old' speakers cannot be assumed to use Basque as solidly as may be expected. According to self-reporting as well as to language-use measurements performed in the street, all-native speaker groups use a great deal of Spanish in conversation, even in Basque-dominant environments. This shows that they are developing language-use and language-choice patterns which are different to those of the previous generation. It should also be noted that they themselves may be providing the language-use model that is so often criticized in the young.

Another interesting challenge that affects language use, but also the consolidation of Basque society, is the integration of new and native speakers. At the moment, the two general types of speakers —new and 'old'— show differences on a number of accounts: attachment to the languages of their repertoire (Spanish/French and Basque), language attitudes, identity, language variety spoken, or their self-image as rightful speakers of Basque, to mention just a few.

The integration of native and new speakers is more complex than it may seem, and it is not always easy for old speakers, the guardians of the minority language, to accept newcomers. Conversely, new speakers can view themselves as not rightful speakers of the language and may even feel uncomfortable or inadequate when interacting with old speakers. This is an under-researched area but there is some evidence which suggests that the linguistic variety spoken by the new and old speakers of minority languages is undergoing a process of revitalization which may play an important part in the speaker's self-concept. In some cases, the linguistic variety has been reported to work as a real barrier. In the Basque context, old speakers have traditionally spoken one of the vernacular dialects, while new speakers have learned the standard *batua*. The images and prestige associated with *batua* and the vernacular dialects have undergone some changes: at first, when the standard *batua* started to be generalized in its use in education, media, and literary production, native speakers of Basque who had not had a chance to develop their language and who were illiterate in it, felt that their variety was sub-standard and uneducated. Luckily, the generalization of Basque in education, as well as the revival of dialectal pride in the last 10 years, is clearly doing away with these attitudes. On the other hand, the standard *batua* is sometimes perceived as too artificial.

We have used the linguistic variety to illustrate some of the differences between 'old' and new Basque speakers, an area that still needs to be researched. However, the integration of the different profiles of speakers is important for a language in the situation of Basque. Old speakers have an important role in mak-

ing new speakers feel comfortable, adequate and rightful speakers, and in helping them to develop solid use patterns and to achieve greater integration into Basque culture.

1.5.2. Fostering affective links with Basque

As discussed in the previous sections, the Basque language policy has increased the instrumental value and the prestige of Basque, as well as the presence of the language in public life (mostly in the BAC). Basque has become an asset for a number of jobs. This newly gained instrumental value also explains why many parents who are not particularly close to Basque nevertheless choose Basque educational programs for their children. For a language with a history of minorization like Basque, this is very positive but poses a new challenge: whilst all the different motivations that act in favor of Basque are legitimate, the challenge is for these instrumental motivations coming *from the outside* to be transformed into motivations that come *from the inside*, *from the heart*, based on a real feeling of attachment to the language and from valuing being part of a bilingual and bicultural community. This is important because it is these kinds of motivations which ultimately ensure the loyalty and commitment to the language that may guarantee its maintenance, even when or if the political situation changes and the pressures from outside disappear.

How can the transition to these motivations *from within* be fostered? A great number of initiatives, both at local and autonomous community level, are being implemented to enable a greater understanding of Basque culture and positive attitudes towards Basque language and bilingualism.

We will now turn to other types of challenges, challenges which all languages, especially minority or small languages, share in a globalized world: among them, the impact of newcomers or immigrants, and new technologies.

1.5.3. Welcoming immigrants to Basque

Immigration in the Basque territories is recent and by no means as extensive as in other Spanish autonomous communities, but it is rapidly increasing. This is a challenge for Basque: newcomers, from a great variety of linguistic and cultural backgrounds, many of whom are speakers of very stigmatized languages (e.g. Quechuan, Aymaran, Tamazigh) arrive in the Basque Country with the expectation of integrating into the dominant language, that is, Spanish; many may not even be aware of the existence of Basque. It is very important that newcomers

understand that they come to bilingual territories and that they perceive learning Basque to be an advantage for integration, if not for themselves, for their children. The experience so far in Basque-speaking areas is that these children successfully learn Basque at school and in everyday life. But many immigrant parents living in Spanish-speaking areas do not choose the Basque immersion model, thus making it more difficult for their children to learn Basque and become trilingual in childhood.

1.5.4. The presence and development of Basque in the new technologies

In relation to new technologies, Basque needs to both increase its presence within these and increase the creation of products in this language. Being active in this field is important: on the one hand, to prevent another language from occupying yet another space or domain of use; and, on the other hand, because of the image and the prestige new technologies have, especially among the young.

Steps have been taken, and the option to configurate products such as Windows or Google has been available for some time now. There is also important activity in Basque-speaking blogs. However, there is still a long way to go in the whole area of new technologies for communication.

1.6. Conclusion

This chapter has discussed the situation of Basque by exploring some areas that show both improvement and uncertainty, or even cause for concern. The recovery of Basque has been dramatic in only 40 years and this is without any doubt the single most positive cause for satisfaction. But this in turn has produced a great number of situations and groups which are not yet well described and whose development is not easy to predict. Several times in this paper, the need to keep investigating the ever-changing progress of the language and its society has been highlighted. This must be done in order to evaluate change and to implement efficient policies that are sensitive to the groups involved. Likewise, it is crucial to continue working to develop positive attitudes and awareness in the society as a whole, so that a small language like Basque can successfully face the challenges that await it in an increasingly globalized world.

2. Basque Dialects

Mikel Martínez-Areta
(University of the Basque Country, UPV/EHU)

2.1. Introduction*

This chapter has two main purposes. First (Section §2.2), I will briefly describe the intelligibility between contemporary Basque dialects, which ones these are —or at least the main dialectal areas—, and how many Basque dialects there are and have been in the historical period. The latter question is rather complex, and although it has been discussed for decades and even centuries, here it will be addressed only along very general lines. Secondly (Section §2.3), I will discuss how and when Basque dialects may have arisen out of a common language, which dialects were the earliest to branch off from the common trunk, and which dialects are historically closer —in other words, which dialectal nets have been denser or more sparse— as a result of that branching process. This issue has attracted less attention over most of the history of Bascology, but, in recent years, the questions posed by Michelena (1987 [1981a]) have given rise to an unprecedented interest in it.

2.2. The Basque dialects today and in the historical past

2.2.1. How different are Basque dialects?

Differences between Basque dialects are not great. The phonological system, the structure of the grammar and the nuclear lexicon are basically the same, or at

* This chapter has been written within the framework of three research projects: 1) The Research Project "Monumenta Linguae Vasconum III-IV: historia, crítica y edición de textos vascos" [= FFI 2008-04516, FFI2012-37696], led by Joseba Lakarra and funded by the Ministry of Education and Science of the Spanish Government; 2) The Research Project "Historia de la Lengua Vasca y Lingüística Histórico-Comparada" (HLMV-LHC), Ref. GIC 10/83, IT 486-10, led by Joaquín Gorrochategui and funded by the Government of the Basque Autonomous Community; and 3) The Training and Research Unit "Hizkuntzalaritza Teorikoa eta Diakronikoa: Gramatika Unibertsala, Hizkuntza Indoeuroparrak eta Euskara" (UFI 11/14), funded by the University of the Basque Country (UPV/EHU). I would like to thank Joseba Lakarra and Iñaki Camino for their comments on a previous version of the work. They do not necessarily share the views defended here, and all errors are mine.

least very similar, throughout the whole of the Basque-speaking territory. In fact, in most situations two speakers of different dialects have no difficulty in understanding each other, and, as Michelena pointed out on several occasions, the consciousness of sharing a common linguistic pool increases as conversation goes on.

The degree of intelligibility, however, varies according to an ample array of factors. The first and probably most influential of these factors is, obviously, the combination of dialects in contact. While two speakers of Guipuscoan and Northern High Navarrese (e.g. from Ultzama) will usually have no difficulty in maintaining a conversation, *ceteris paribus* two speakers of Biscayan and Souletin —located at the opposite ends of the Basque-speaking territory— will find it more difficult to communicate in Basque. A second factor is the speakers' degree of consciousness of the common linguistic features that they possess. This does not necessarily imply knowledge of the standard language, or *euskara batua*, although, in the last decades, thanks to the promotion of this *koiné*, this is the most straightforward tool that speakers can resort to in order to get around otherwise insurmountable dialectal differences. Very often, the possibility of communicating in the common language relies on the subjective attitude of the two speakers, or at least of one of them, towards that possibility. It is still common, especially among old native speakers, to come across skeptical attitudes about the pertinence of a unified Basque, and a consequent emphasis or exaggeration of the features of their own variety. Obviously, the set of parameters which universally condition communication, such as the age, gender, profession or social background of the speakers concerned, is an element which also plays a role within this variable. A third factor is the administrative division of the Basque Country. This especially affects the lexicon, since continental and peninsular speakers are acquainted with different terms of frequent use in everyday life.

Two of the factors just mentioned are actually sociolinguistic. Lakarra (2011f) has compared the linguistic diversity among South American languages, as described by Cerrón-Palomino (1987), to that of Basque. With regard to that diversity, and specifically to that among Quechuan languages, this author pointed out that:

> "The general impression that one obtains is that, in fact, the issue about the intelligibility between Quechuan languages is more sociocultural than linguistic. The lack of contact between speakers, and the intrusion of Spanish as a mediating language (avoiding any effort to understand and to make oneself understood through the interlocutor's dialect) cause language barriers, not as impassable as was believed, to be magnified. Should these difficulties be overcome, it would be perfectly possible for Quechuan speakers to communicate in a bi-aural way, that is, each using his/her

> own variety" (Cerrón-Palomino 1987: 229 [translation into English mine, MMA]; apud Lakarra 2011f).

This observation can be applied to the intelligibility between Basque dialects, and in fact is to be expected in any diglossic situation in which the dominant language provides a more powerful tool to manage a number of situations, such as for instance administrative issues. This preponderance of the dominant language manifests itself most clearly in written registers, since unlike the vernacular language it has a standardized orthographic system and an elaborate vocabulary for its administrative or bureaucratic lexicon. We have an attestation of the impossibility, or at least the ineffectiveness, of communicating in writing in the form of correspondence between the city councils of Hondarribia (on the G. coast) and Urruña (on the L. coast, close to Hondarribia) in 1680 (*ContTAV*: 132 ff.). The authorities of Hondarribia wrote their first letter in Spanish and the second —now lost— in Basque. To this, the authorities of Urruña answered asking them to switch back to Spanish in the future, as they had difficulties in understanding the letter in Basque.

The awareness of local differences is attested from almost the very beginning of Basque literature. Thus, in the preface of his translation of the New Testament (1571), Leiçarraga states: *(...) batbederak daki Heuskal Herrian kasi etxe batetik berzera-ere minzatzeko manerán zer differentiá eta dibersitatea den* '(...) everybody knows what a great difference and diversity there is in the way of speaking in the Basque Country, almost from one house to the next' [translation mine, MMA] (cf. *EDH*: XXXVI-XXXVIII, for more of such references in Basque literature). In fact, even in the 16th century —and in some respects even today— many valleys saw themselves to some degree as small republics, as Dch. already records in at least two poems (XIV, 2-3: *Garazi-ko herri-a / benedika dadi-la* 'let the land of Cize be blessed'; XV, 2: *O heuskara lauda eza-k Garazi-ko herri-a* 'Oh, Basque language, praise the land of Cize!' [translation mine, MMA]). Their inhabitants considered neighboring varieties to be somewhat alien, and sometimes mocked their accent or their vocabulary.[1] This has

1 Caro-Baroja recalls, citing a reference made by Lope Martínez de Isasti in 1618, that, at the time of the witchcraft trials in Navarre and neighboring regions, for Guipuscoans, the continental and Navarrese Basques were potential suspects of witchcraft due to their rare accent, and he himself reports that in Bera he had met a woman who got excited when she heard anybody speak the Basque of other regions, because she said that they were witches or warlocks (apud Pagola 1991: 77). Without reaching such extremes, it is common for the variety of a certain valley or place to sound particularly peculiar to the Basques of the surrounding areas, as is the case with the Basque of Arantza, which not by coincidence is orographically delimited. Here, the aphereses and syncopes typical of the whole area of Bortziriak-Cinco Villas (§2.2.4) are extreme, and people from the

also fostered linguistic atomization. Of course, interregional relationships were warranted by commerce (by means of fairs, markets, etc.), or by treaties on the distribution of grazing lands among shepherds (Sp. *facerías*), but other administrative issues had to be solved in institutions of a higher level, where the only language employed was Romance.

2.2.2. What are the Basque dialects?

In this subsection, I am going to give a succinct enumeration of the Basque dialects, a very brief description of their extension and recent sociolinguistic status, and a summary of the most distinctive features of each one (as well as their importance for diachronic research), in the form of a quick tour throughout the Basque-speaking territory.[2] I will basically follow the classification made by Prince Louis-Lucien Bonaparte in the 19th century (cf. §3.4.3), which has been traditional until recently, and is still standard, with greater or lesser additions or retouchings depending on the purpose aimed at, in works of reference for philologists such as the *OEH*. The Alavese will be added to that group of 8 dialects, but some others such as R and Sal, which are handled separately in Michelena's *FHV*, will be treated as subdialects of S and East.-LN, following Bonaparte's division. The sociolinguistic data are taken from *EDH*, and belong to the years 1970-72 (the source being Yrizar 1973, reedited in Yrizar 1981: 171-220), or to the year 1981 in the case of Guipuscoan. I think that this is more appropriate than giving today's data as indicators of the real number of each dialect, since these can be distorted by the irruption of SB or *batua* in the last few decades.

Zuazo's more recent alternative classification is explicitly based on innovations (1998b: 192), and hence is less useful for the diachronist. In fact, the election of features which will be mentioned for each dialect is extremely selective, and subordinated to the diachronic enterprise of the volume. Thus, some synchronically important phonetic issues will be left out, as will the lexicon. I will

---

neighboring villages say that they speak in a rare manner. A person from Etxalar reported: "Those of Arantza speak as if they were in a hurry, they drop more than half of what they say" (Zelaieta 2008: 124-127 [translation mine, MMA, from a testimony recorded by Zelaieta]). The study of the impressions that (non-linguist) speakers of each variety have about the other varieties, called *perceptual dialectology* in some circles (Basq. *oharmenezko dialektologia*), is a useful tool according to Basque dialectologists, which in any case must be handled with care (Camino 1997: 45-48, 263-275).

2 Cultural and anthropological aspects will be briefly touched upon when they are relevant to better look into the situation of a dialect, but historical aspects will be skipped. A succint history of the Basque Country, which pays attention to linguistically relevant issues, is provided in Trask (*HB*: 7-35).

also skip micro-geographic nuances, and some other details which would make the analysis more intricate.

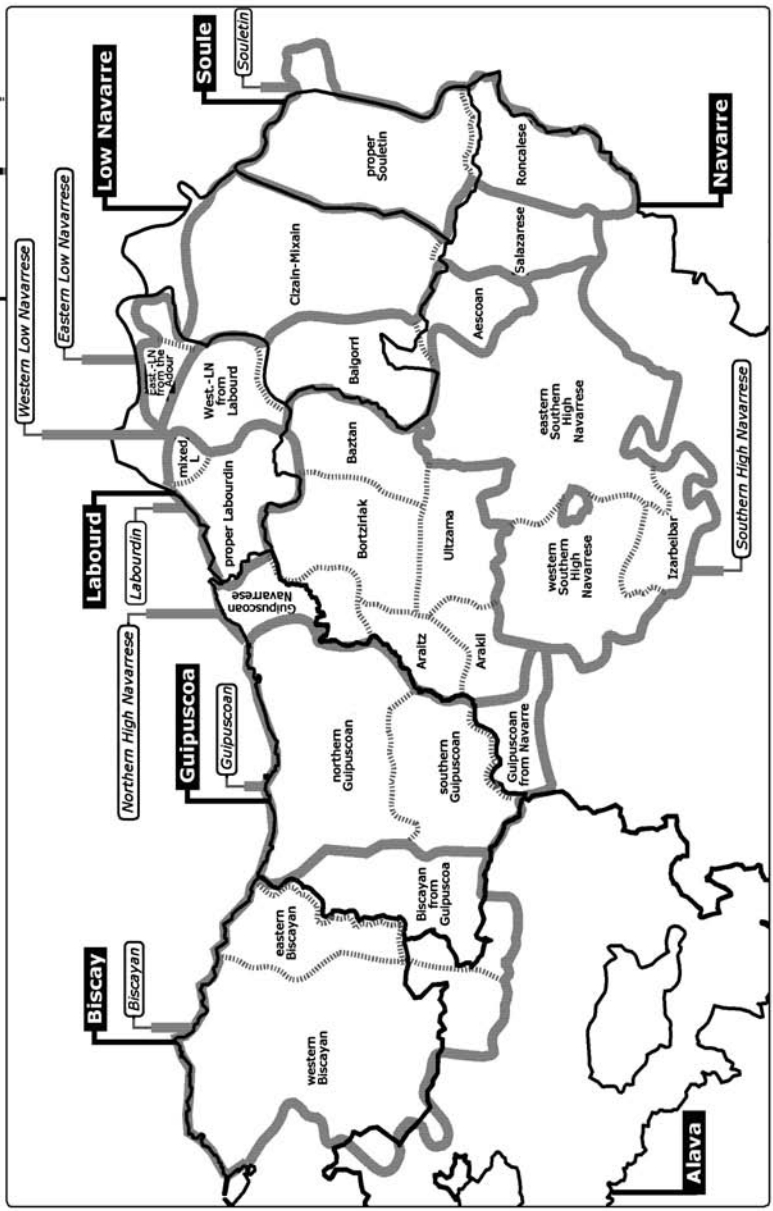
It is important to note that, within the domain that I am going to describe as belonging to a Basque dialect, not everything is entirely Basque-speaking, far less so today than it was in 1970-72, and far less in 1970-72 than it was in Prince Bonaparte's times. In each place within that area, the proportion of Basque speakers will be inversely proportional to the degree of urbanization of that place, and, in general, it will be lower in regions bordering with monolingual (Spanish or French) territories. Furthermore, due to the diglossic situation described in §1.2, even in the most Basque-speaking areas bilingualism and code-switching is frequent. Conversely, there may also be speakers of Basque outside the outline of the Basque language (e.g. in Vitoria-Gasteiz), but these will almost always be *euskaldun berris* (i.e. 'new Basque'), who speak Standard Basque.

Having said that, we can now look at Maps 2.1, 2.2 and 2.3, where the dialects, subdialects and varieties established by Bonaparte in his last classification are represented upon the administrative borders of the Basque provinces (Alavese is left out as a dialect, although it is described in §2.2.2.3).

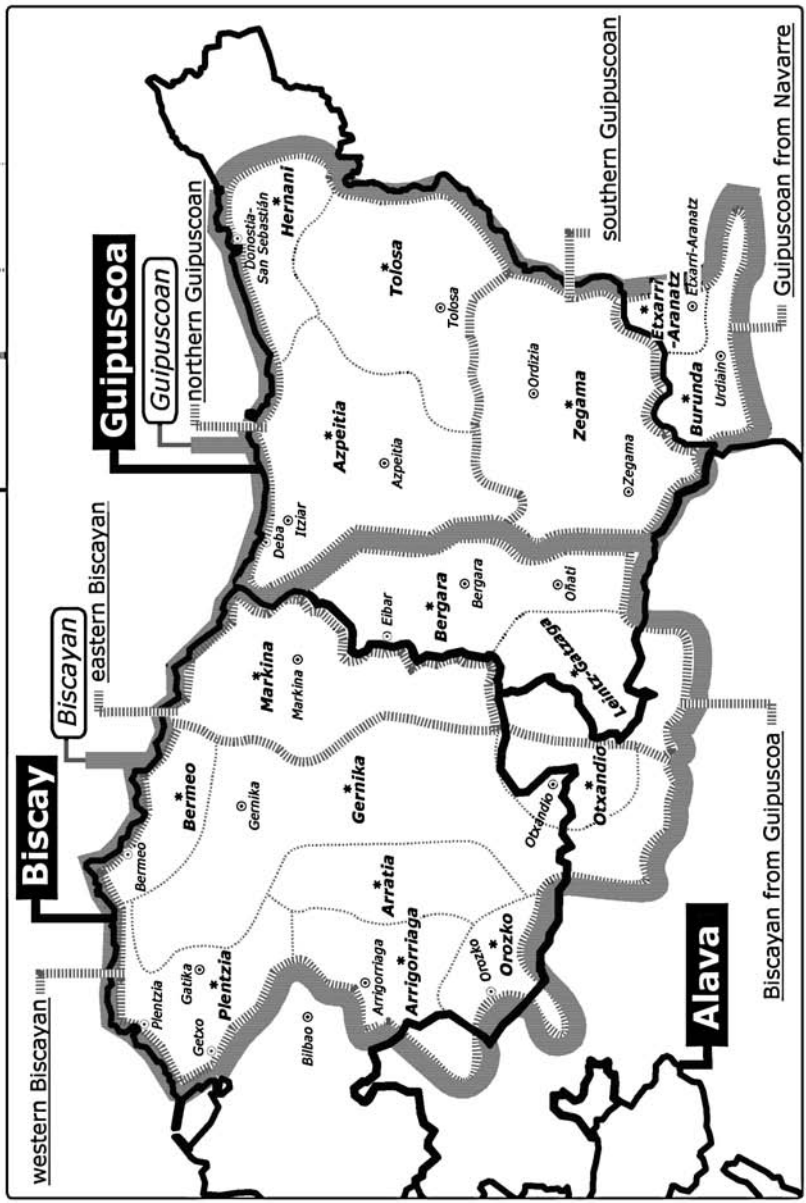
2.2.2.1. Biscayan

Biscayan is the dialect spoken in the two central and eastern thirds of Biscay, and in a western strip of the administrative territory (*herrialde* in SB, although this term is very recent) of Guipuscoa, which roughly corresponds to the Deba Valley. In Bonaparte's fourth classification, it was subdivided into three subdialects: 1) Western Biscayan (with the varieties of Gernika, Bermeo, Arratia, Orozko, Plentzia, Arrigorriaga and Otxandio); 2) Eastern Biscayan (with Markina as the only variety); and 3) Biscayan from Guipuscoa (with the varieties of Bergara and Leintz-Gatzaga).

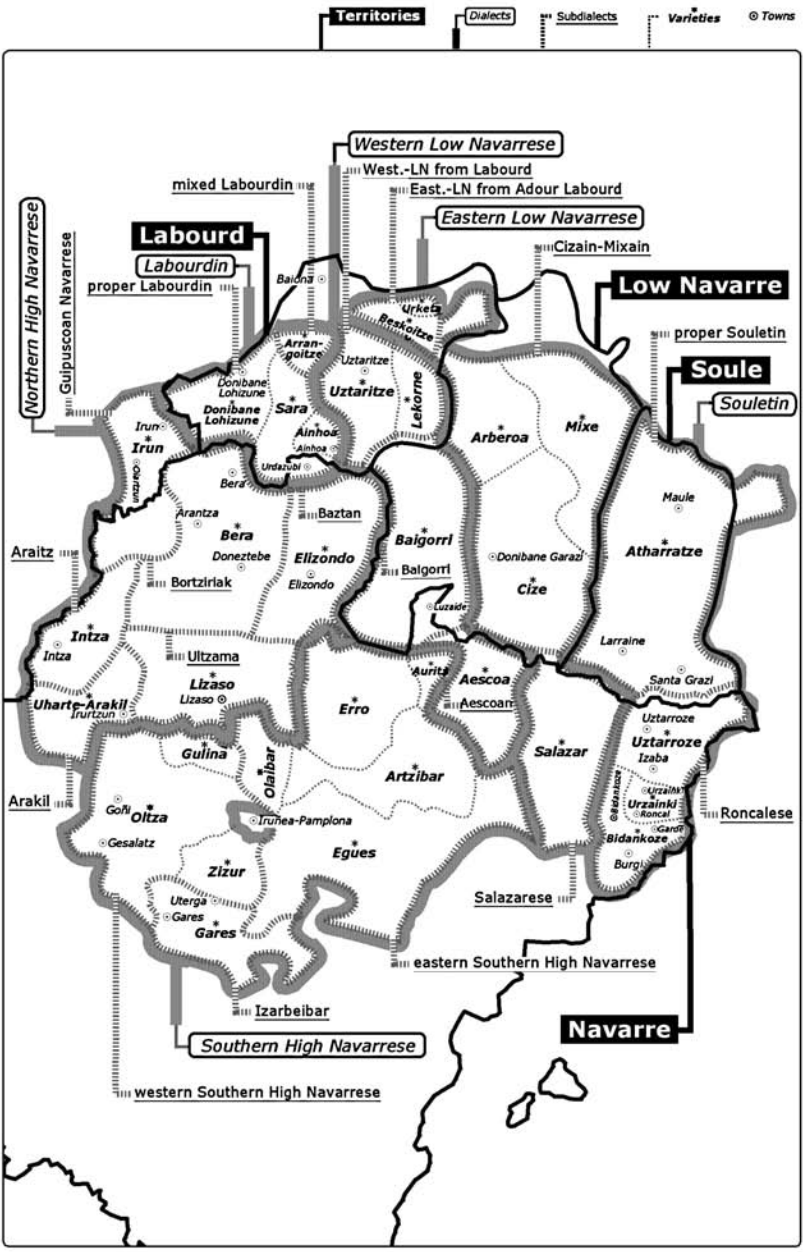
In a census of 1970 (apud *EDH*: 12), the number of speakers of Basque in all those varieties (i.e. of Biscayan) was 193,494 (out of a total number of 477,652 inhabitants). Bilbao is the capital of Biscay, and in many aspects it is also the flagship of the Basque Country, although the capital of the Basque Autonomous Community is Vitoria-Gasteiz. Bilbao is mainly Spanish-speaking today. Basque was widely used there in the 16th and 17th centuries, although always in competition with Spanish or Castilian. The impact of the industrialization of the 19th century reduced the presence of Basque in Bilbao and the left side of the Nervión River to a minimum. To the west of Bilbao, there is toponymic evidence that Basque was spoken, at least in the Middle Ages, in a strip of the area of Encartaciones, and to the southwest in the Ayala Valley (Alava).



Map 2.1. Divisions of territories, dialects and sub-dialects of the Basque Country



Map 2.2. Varieties and towns in the west



Map 2.3. Varieties and towns in the east

Nowadays, schooling in Basque has enabled most children to be at least passive or potential speakers of Basque, but the practice of that knowledge highly depends on the environment, the family, etc. Despite this, the proportion of speakers of Basque has risen with respect to 1970, but at the expense of prioritizing Standard Basque. Some varieties or subvarieties of B which are extremely interesting for diachronic research, such as those of Getxo or Plentzia, have very few speakers.

Biscayan is a highly idiosyncratic dialect, with many features which differentiate it from the other dialects. It is also relatively well documented, and although B does not acquire prestige as a literary dialect until the 19th century, from the end of the 16th century onwards we have a number of texts which can even, to a certain extent, permit us to outline the diachronic development of a number of geographically scattered varieties (§3.4.2).

A phonetic feature characteristic of modern B is the merger of the apico-alveolar and dorso-alveolar sibilant fricatives *s* /s̻/ and *z* /s̺/ into the apico-alveolar *s* /s̻/, and the merger of the apico-alveolar and dorso-alveolar sibilant affricates *ts* /t̻s̻/ and *tz* /t̺s̺/ into the dorso-alveolar *tz* /t̺s̺/. This merger is already accomplished, or nearly accomplished, in Mic. and VJ, texts of western Biscay of the 17th century, and consequently it appears to have spread from west to east, and from larger towns to villages (*FHV*: 282-283, Ulibarri 2010: 58-59).

In the noun morphology, the B. sg. paradigm undergoes a morpho-phonological process whereby, if the article is attached to an *a*-stem, the group *-a-a* is dissimilated into *-ea* (*neska* 'girl' → *neskea* 'the girl'). This takes place in the abs., erg., gen. and dat., but not in the ines. nor in the plural.[3] This feature is shared by the G. areas of the Urola and Goierri Valleys and by Alavese, and it even penetrates into the Navarrese Burunda. This situation is attested by the 16th century, but is actually the starting point for a push-chain process, whereby, as *-a-a* develops into *-ea* (or even into *-ia* or *-ie*, with assimilation), *-e-a* in turn develops[4] into *-ia* (or *-ie*), *-i-a* into *-iy/xa* (or *-iy/xe*; *x* being /ʃ/, which can also develop into /ʒ/), *-o-a* into *-ua* (or *-ue*), and *-u-a* into *-u(b)a* (or *-u(b)e*). The result is a plethora of possible outcomes, the development of which is visible from the

3 However, the ines. and the pl. are archaic in Old B and Al in that the etymological junctures ines. *-a-an* and pl. *-a-ak* have kept the geminate vowel, as in Laz. *domekaan* 'on Sunday' and Laz. *berbaak* 'the words', whereas in other dialects both *a*-s have merged into one. In fact, this feature has survived until recently in at least some neighborhoods of Oñati and Aretxabaleta, in the High Deba (cf. Zuazo 2006: 163-164).

4 Or rather it can develop, as there are varieties which do not go beyond the change *-a-a* > *-ea*. The process is ideally sketched, and mentions only some possibilities (cf. the maps in *EDH*: 22-27).

16th century onwards, and characterizes the extreme linguistic diversification of the western area.

One of the most distinctive features of B is the way of forming the soc.: sg. *-a-ga-z*, pl. *-akaz* (< **-ak-ga-z*). This also appears in Al (not exclusively, and in fact Lcc. always or almost always has *-(ar)ekin*; Zuazo 1998a: 148), but in the rest of the dialects this case is formed by *-(ar)e-kin* (or *-(ar)e-ki(la)*), also used in east. B. regions, and more in the pl. than in the sg. (Zuazo 2006: 75, Mugarza 2006: 90). *-(ar)e-kin* is a secondary case formed upon the old form of the genitive, and hence comes from a periphrasis. One possibility is that the distinction between B and the other dialects on this point came about because each group morphologized a form from a common pool of possibilities. Nevertheless, it seems more probable that *-(ar)e-ki(de)n* existed in the common language, and that B created a new sociative in the dialectal period (§6.5.2.1).

Old B and Al —and to a lesser extent Old G— have an abl. *-rean* (*etxe-rean* 'from the house'). This, however, appears to be a western innovation which did not succeed and finally disappeared (§6.4.2.3). Its rise must be related to the fact that *-ti* was only pros. and did not take over any abl. function, unlike in the east. The pros. is, by the way, an independent case in old west. texts.

The plural of 1st and 2nd grade demonstratives is another feature which separates B from the rest of the dialects (also from Al, at least from Laz.). The distinction points to: B. *ón-eek* / cent. east. *hau-ek* 'these', B. *órr-eek* / cent. east. *hori-ek*. These appear to be two different ways of building the plural, perhaps at some point of the dialectal fragmentation, but pointing to the incompleteness of its formation in the common language (§7.2.6).

B has a number of distinctive features in the verbal morphology, shared or not by adjacent areas (especially Al). In the non-finite verb, the ipfv. allomorph *-te* appears in more contexts than in the other dialects (*ekar-ten* 'bring (ipfv.)', *ibil-ten* 'walk (ipfv.)', *jo-ten* 'hit (ipfv.)', vs. cent. *ekar-tzen*, *ibil-tzen*, *jo-tzen*), although some of these phenomena have parallels in the east (§8.3.2.2.1), and *-(k)eta*, with a parallel in the ipfv. R. *-(e)tan*, and which appears as a derivational suffix in all dialects, has succeeded as another ipfv. allomorph, with Al being the possible irradiating point of that spread (§8.3.2.2.3 and §8.3.2.2.4). An unsuccessful possibility, however, was the ipfv. allomorph *-(t)zaite* (§8.3.2.2.5). All these appear to be diverse grammaticalization processes from what was derivational or lexical in a common pool.

In the finite verbal morphology, B and Al also share some features. One is the lack of *z-* in past forms in which the word-initial morpheme is of 3rd p. (B. *ebilen* / cent. east. *zebilen* '(s)he walked'), except with the verb *izan* (*zan* '(s)he was', *zirean* 'they were'), sbjv. (< aor.) Nor auxiliaries (*zidin* '~ (s)he was', *zitean* '~ they were'), sbjv. (< aor.) Nor-Nork auxiliaries (*zezan* '~ (s)he had', *zitzan*

'~ (s)he had (several things)'), Nor-Nork auxiliaries with abs. pl. (*zituan* '(s)he had (several things)'), and some others.[5] There is as yet no consensus as to whether B is archaic or innovating on this point (cf. §9.4.1.1.2). Another important feature of B is the vocalism of its bipersonal forms of the Nor-Nork (abs.-erg.) type (§9.5): B. *dot* / G. *det* / east. *dut* 'I have', *dau* / G. east. *du* '(s)he has', B. *neban* / G. east. *nuen* 'I had'. These, however, only represent different phonetic solutions of morphological forms which were already set up in the common language: **da-du-da*, **da-du*, **ne-dun(-a/e)-n*.

By contrast, in the auxiliaries of the Nor-Nori (abs.-dat.) type with 3rd p. abs., B appears to have opted for another root: *ja-t* / G. east. *zai-t* 'it is to me'. The root of Nor-Nori-Nork (abs.-dat.-erg.) auxiliaries is also different from the rest of the dialects: *d-euts-o* / G. HN. *d-i-o* / east. *d-erau-ko* '(s)he has to him/her'.[6] Since *eutsi* 'to cling to' is a verb in the lexicon of peninsular dialects, here again we might have the grammaticalization of an element which was lexical in the common language (§9.5.4).

More intricate is the existence of pot., sbjv. and imp. finite forms in B built upon the root **gin* (*e-gin* 'to do, to make'), as opposed to **zan* and **iro* in the rest of the dialects (§9.4.2). The root **gin* in such contexts also appears in Al and in a large part of G,[7] and on the other hand **zan* also appears in Old B and Al, but never in Nor-Nori-Nork auxiliaries (with one exception in Barr.). After a detailed analysis of the texts which have pot., sbjv. and imp. verbal forms with

5 This varies throughout the west. territory, as if we move eastwards from west. Biscay, we progressively find forms like *zegian* '~ (s)he had', *zeban* '(s)he had', or even Nor-Nori-Nork auxiliaries like *zeustan* '(s)he had to me' (in B. regions from Eibar to Bermeo; cf. Michelena 1987 [1981a]: 43).

6 This tripartite division is approximate, and reflects today's situation, if any. Old texts offer us, as usual, puzzling data with regard to Nor-Nori-Nork roots. For the situation in the continent and Navarre, cf. §2.2.2.6. In addition, Old G has some Nor-Nori-Nork auxiliaries, with abs. pl., where the root appears to be *-erau-*: Och.A. *d-irau-z-ku-n* '(s)he has to us (several things)' (SB. *dizkigun*), *z-irau-z-ku-n* '(s)he had to us (several things)' (SB. *zizkigun*), *d-irau-tz-a* '(s)he has to him/her (several things)' (SB. *dizkio*) (see Lakarra 1985 and 2011f for more of such cases). Borja Ariztimuño suggests (apud Lakarra 2011f) that the rare *diozka* '(s)he has to him/her (several things)', still present today in some G. varieties, might come from *diuzka* < *diauzka* < *dirauzka*, which would imply that the root *-erau-* still survives in G, albeit hidden. The root *-erau-* also shows up in Laz., and not only with abs. pl.: *ez deraus-ta-e beste-rik atera-ko* 'they will not take me anything out'. Note that in the nearby Sakana there are also Nor-Nori-Nork forms with (phonetic developments of) the root *-erau-* (Zuazo 2008: 201).

7 In Old G. texts, the tendency is: 1) in Nor-Nork auxiliaries with abs. sg., the root can be **zan* or **gin*, apparently indistinctively; 2) in Nor-Nork auxiliaries with abs. pl., the root is **zan*; 3) in Nor-Nori-Nork auxiliaries, the root is **gin* (Lakarra 1996a: 181-185).

the root **zan* and those which do not, Lakarra (1996a: 184-185) suggests that the total substitution of **zan* through **gin* in B spread from the south (of Biscay) to the north and east (cf. also Mounole 2007). It is possible that Old G, and partially Old B, maintain the state of affairs of the common language, with **zan* being general in Nor-Nork auxiliaries with abs. pl., and **gin* being general in Nor-Nori-Nork auxiliaries. The center and east generalized **zan* (in competence with **iro*) for every context, whereas B ended up generalizing **gin*.[8]

Another element which characterizes B is the productivity reached by the abs. pluralizer *-z*, located at the end of the verb (but immediately before *-n* in past forms). Thus, we can find, depending on the variety, Nor-Nori auxiliaries like *jata-z* (vs. G. L. *zai-zki-t*, S. *zai-z-t*, etc. 'they are to me'), Nor-Nork auxiliaries like *doda-z* (vs. comm. *d-it-ut* 'I have (several things)', with the abs. pluralizer *-it-* of forms with **edun*), *dau-z* (vs. comm. *d-it-u* '(s)he has (several things)'), *eba-z-an* (vs. comm. *z-it-ue/an* '(s)he had (several things)'), or synthetic verbs like *dakarda-z* (vs. G. *dakar-zki-t*, L. *dakar-tza-t* 'I bring (several things)'), etc. This is an innovation, consisting in the generalization of one of the several possibilities for pluralizing the abs. in the common language (cf. §9.3.2). In Nor-Nork auxiliaries and in some Nor ones, the spreading process of such forms is observable throughout the written tradition, since in the earliest texts only forms like *d-it-u*, *z-it-uan*, or pleonastic forms like *d-it-u-z*, *z-it-u-z-an*, appear in B.[9]

---

8 Perhaps Al also shows traces of this situation. Zuazo (1998a: 158) states that, between **zan* and **gin*, Al opted for the latter. Of the three exceptions that he presents, in the Alavese author García de Albeniz, one of them is *barkatu ditxadala* 'that I forget them', a Nor-Nork auxiliary with abs. pl. In Laz., **gin* is more frequent than **zan* in pot., sbjv. and imp. auxiliaries, and, more significantly, **zan* never or hardly ever appears in Nor-Nori-Nork auxiliaries.

9 Things are different, however, for Nor-Nori and Nor-Nori-Nork auxiliaries, in which the pluralizer is always *-z* from the beginning of the tradition: *jata-z* 'they are to me', *deusku-z* '(s)he has to us (several things)'. Perhaps the Nor-Nori-Nork auxiliaries were a starting point for the spread of the pluralizer *-z*. In these, the presence of the root *-euts-*, a sibilant-ending root, may have made the suffixation of the pluralizer *-z-* unfunctional, as a form like ***deus-z-ku* (abs. pl.) would hardly oppose to *deus-ku* (abs. sg.), for phonetic reasons. A solution could be to bring the pluralizer to the end of the word.
While this abs. pluralizer has reached such a high productivity only in the west, it appears that the possibility of using that abs. pluralizer was latent from the common language, as in the east we also find forms like *gu-k ekarr-i gindu-z-en* 'we brought them' (SB. *gen-it-uen*), uttered by a Labourdin gypsy in a trial for theft in Navarre in 1597 (*ContTAV*: 111), or even Dch. *Ieinko-ak egin zindu-z-en* 'God made you' (SB. *zin-t-uen*).

Old B and Al present verbal personal indexes of 2nd p. pl., like abs. *za-* and erg. *-zu* (Laz. *zuek jakin-go do-zu* 'you will know'), which have been reinforced in the rest of the dialects (→ *du-zue*), in order to mark their plurality and distinguish them from the indexes of 2nd p. sg. (of respect). This is obviously an archaism (cf. §7.3.1, §9.3.2). By contrast, the past verbal forms without *-n* attested in northwest. B, as in RS. *ezer ez nekarre* 'I didn't bring anything' (*nekarren* in the rest), must be innovations (cf. Lakarra 1996a: 166-167), despite parallel forms in South.-HN and adjacent regions.

2.2.2.2. Guipuscoan

Guipuscoan is a dialect with well-defined features, but not as idiosyncratic as B. It is perhaps the most vigorous dialect, as it is the only one which is surrounded by Basque-speaking areas on almost all sides. Furthermore, although in urban centers Spanish is the dominant language, Basque is spoken throughout the whole territory. Thus, although in the capital Donostia-San Sebastián it is minoritary, it is considerably more frequently employed than in the Biscayan capital Bilbao. According to Bonaparte's fourth classification, G is divided into three subdialects: 1) Northern Guipuscoan (with the varieties of Hernani, Tolosa and Azpeitia); 2) Southern Guipuscoan (with the variety of Zegama); and 3) Guipuscoan from Navarre (with the varieties of Burunda and Etxarri-Aranatz).[10]

According to the census of 1981 (apud *EDH*: 87-89), the number of speakers of all Guipuscoan municipalities belonging to the G. dialect (but leaving the Burunda and Etxarri-Aranatz aside) was 197,020 out of a total number of 486,098 inhabitants.

As a literary dialect, Guipuscoan did not emerge until the 18th century, although we have a number of texts in G belonging to the 16th and 17th century (§3.4.1).

10 The ascription of these varieties to G has been much discussed. Apparently, their inclusion within G comes from the fact that the translators from Urdiain (Burunda) and Etxarri-Aranatz that Bonaparte entrusted with the translations of texts misunderstood the Prince's intention (to reflect the local speech), and instead used the prestigious Basque of the time: the Guipuscoan of the Beterri area (between Donostia-San Sebastián and Tolosa). This misconception was already noticed by Damaso Inza in 1922 (apud Zuazo 2010b: 191-195). In fact, later researchers have considered the classification of Burundan Basque as G to be too simple. Yrizar considers it to be of uncertain classification, since it contains B., G. and HN. elements. Zuazo considers it to be a variety of transition, whereas the variety of Etxarri-Aranatz, and the one from there to the east, as well as that of Ergoiena, are considered to be HN (apud ibid.).

Linguistically, as we have seen in §2.2.2.1, a considerable number of G. features represent an abrupt break away from B (and usually also Al) and align with the dialects to the east, whereas some others represent a continuity with the west,[11] which in the case of the erg. pl. *-ak* and the dat. *-ai* also comprehends the western Navarrese territory.

Perhaps the most typical and distinctive feature of G is the vocalism of the Nor-Nork auxiliaries of the present tense: *izan det* 'I have had', *izan degu* 'we have had', etc. (but *izan du* '(s)he has had').[12] In addition, some northeastern G. varieties pluralize the 2nd p. pl. erg. morpheme *dezu* 'you (sg.) have' not by *-e*, but by *-te* (*dezute* 'you (pl.) have'), which sounds quite Guipuscoan, although it also turns up in some North.-HN. varieties (here *duzute*). The following Nor auxiliaries are also typical of G (although they also penetrate into some North.-HN. areas): *zera* 'you are' (as opposed to B. *zara*, North.-HN. *zara*, L. *zare*), *gera* 'we are', etc. These appear to be the result of a vowel dissimilation *a-a* > *e-a*, although it is also conceivable that they come from a vowel metathesis of *za-re-*, etc., which in turn would come from a reanalysis of the suffixed forms like *zare-la* 'that you (sg.) are' (< **zara-ela*), etc., as Ariztimuño interprets (§9.8.1). Be it as it may, they appear from the earliest G. texts.

As a final verbal feature, I will mention that the abs. pluralizing morpheme *-zki-*, which is a central innovation created from a reanalysis (*HB*: 222), reaches its highest productivity in G, since there it is not only exclusive in Nor-Nori-Nork auxiliaries (in most central varieties), but has also spread to synthetic verbs such as *darama-zki* '(s)he carries (several things)' (as opposed to *darama-tza* to the east), and *dakar-zki* '(s)he brings (several things)' (as opposed to *dakar-tza* to the east).[13] The pluralizer *-zki-* also spread to allocutive forms such as *di-zki-a/na-t* 'I have it, man/woman' (as opposed to SB. *d-it-i-a-t*).

Another form which sounds quite Guipuscoan, although in fact it is shared by North.-HN. and some South.-HN varieties, is *nere* 'my', an innovation created in the historical period (§7.3.2).

---

11 And there are also forms which combine both possibilities, such as Och.A. (Ordizia) *zeban* '(s)he had', which has the vocalism of past Nor-Nork auxiliaries of the west but the *z-* of the center and east; or even *zuan* from Itziar (Yrizar 1991-2008), which also has cent. east. vocalism, but still preserves the *-a-* as the binding vowel (as against L. HN. *zu-e-n*).

12 There are, however, G. varieties in which we find forms of transition between B. *dau* / G. east. *du*, such as *deu* ~ *do* in Zegama, or *du* ~ *deu* in Azpeitia (Yrizar 1991-2008).

13 At least the first of these, *darama-zki*, stretches to some Navarrese territories. In Baztan, both *da(ra)ma-zki* and *da(ra)ma-tza* can be heard (Salaburu & Lakar 2005: 128).

2.2.2.3. Alavese

This dialect no longer exists today, except in the valley of Aramaio (the northern region of Alava which sticks into territories of Biscay and Guipuscoa) and some surrounding neighborhoods. In the 16th century, however, it was spoken perhaps in the entire northern half up to Vitoria-Gasteiz —where Lcc. most probably obtained at least two informants for his dictionary—, and probably even further south, albeit in competition with Castilian. Since then, it has receded, leaving only a few records in the following centuries, such as Gamiz's poems of the 18th century. By the 19th c. and the beginning of the 20th c. it was spoken only in a thin northern strip of Alava (apart from Aramaio), where ethno-linguistic material such as tales were gathered by Azkue and some other researchers. But, according to Bonaparte, this was not enough for it to be considered an independent dialect, so he included those neighborhoods within B.

Modern philological practice, however, usually considers it to be a dialect, among other reasons because it is more practical for comparative argumentation. In fact, the 16th century material (Bt., Lcc., and since 2004 Laz.) is extremely useful because it provides records which sometimes align with the west, less frequently with the continent and the east in general, sometimes with none, and, still others, offer unexpected elements. In particular, Laz. is an excellent source, since, in addition to providing quite a large corpus, it occasionally reveals archaisms which are only, or almost only, paralleled in eastern regions, and hence they can be postulated for the common language.

All this information can be complemented by toponymic material. In fact, Alavese toponymy is particularly rich in Basque elements, and these give us important clues not only about the lexicon but also e.g. about morpho-phonological processes in Alavese for phases prior to the written tradition. In eastern areas, we even find Satznamen such as *Goicoeliseadana* 'which is the church of the top'. Basque toponymy also reaches the High Rioja and northeastern regions of Burgos. An extremely useful source in this respect is the *Reja de San Millán* (*CSM*, 1025), which provides a list of all the villages which paid a tax to the Monastery of San Millán in the High Rioja, many of those toponyms appearing in an archaic form. All this evidence bears out the view that Al was an extremely dynamic dialect throughout the Middle Ages, and Zuazo (2010a) has argued that Vitoria-Gasteiz was most probably an irradiating focus of innovations (§2.3.4).

In general, Al aligns linguistically with B, but the sources range from the southeastern Gam. to the tales gathered in the north (Zuazo 1998a), and hence there is diversity. A phonetic difference of Al with respect to B is that the former distinguishes old *-ain*, *-uin* from old *-an*, *-un*, whereas the latter has merged old *-ain*, *-an* into *-an* and old *-uin*, *-un* into *-un* (*FHV*: 139): RS. Mic. Cap. VJ. *gan*

'top' vs. Lcc. *gainean*, *gañean* 'on the top', RS. Cap. *mun* vs. Lcc. *muin* 'kiss' (cf. Zuazo 1998a: 136-137 for details). The western morpho-phonological process *-a-a* > *-ea*, etc. is realized in Al. However, especially in the middle of Alava there are some toponyms which attest a peculiar development of the juncture formed by the derivational suffix of abundance (or place) *-(t)za* + the article *-a* (Zuazo 1998a: 136): *Bustinzaya* (*buztin* 'clay'), *Elorzaya* (*elorri* 'hawthorn'), *Loizaya* (*lohi* 'mud'), etc. This development is also attested in some words gathered in Aramaio and the High Deba, as in *bedártzaia* '(plenty/place of) grass' (*belar* 'grass') (apud ibid.: 136).

Actually, Al reveals a special linguistic proximity to the High Deba (which includes the territory corresponding to the variety of Leintz-Gatzaga), as one would expect from the commercial route which linked Vitoria-Gasteiz to the coast (to the town of Deba) along the valley.[14] One linguistic feature shared by Al, the High Deba and some other areas of southeast. B and south. G is the lack of abs. pl. agreement marker in the corresponding verb, when the direct object is plural and definite, as in Laz. *Neure begi-ok giadu-ko dot* 'I will guide my eyes' (Laz. *d-it-ut*, with abs. pl. marker, also turns up, but seldom and unsystematically). Another one present in an area covering Al and some areas of the High Deba is the removal of the pluralizer *-it-* of Nor-Nork auxiliaries when the abs. is 2nd p. sg. (of respect) or 1st p. pl., as in *serbidu-ko zaut* 'I will serve you (sg.)', where we would find *zaitut* in other dialects (in fact, Laz. also has *za-it-u-t*, and even one case of the pleonastic *za-it-u-da-z*). In this *zaut*, the pluralizer *-it-* has been removed because *za-* is lexically plural in itself. Hence, this must be an innovation, and so must the elimination of the abs. (definite) pl. agreement.

2.2.2.4. Northern High Navarrese

If we now move from the Alavese Plain into Navarre in the direction of Pamplona-Iruñea, we have to pass through a corridor called the Sakana (from *sakan* 'hollow'). The first stretch, leaving the Urbasa Mountain Range on the right, called the Burunda, was included within the G. dialect by Bonaparte's definitive classification, and has been commented on in §2.2.2.2.[15] The second stretch,

14 There was another route, to Bermeo (on the B. coast) through Durango, but its course was longer and craggier. It is also worth recalling that Oñati (the easternmost town in the High Deba) was not annexed by Guipuscoa until the mid-19th century (Zuazo 2006: 154).

15 To the south of the Urbasa Mountain Range, there are two valleys, Amescoa and Lana, in which Basque was spoken in the Middle Ages, and probably later or much later. Michelena (*TAV*: 121) points to the possibility —among others— that Martin Portal, the author of the poem which was awarded a prize in 1610 in Pamplona-Iruñea, might come from there. Otherwise, the toponymy of this region was studied by Salaberri-Zaratiegi

from Etxarri-Aranatz onwards and already within North.-HN, runs between the Andia Mountain Range on the right and the Aralar Mountain Range on the left. At the end of the corridor we arrive in Irurtzun, and if we turn left there (that is, to the north), we enter the very heart of the Northern High Navarrese territory. We have three basic areas here: the Ultzama Valley and its surroundings (Basaburua, Imotz, Atetz, Lantz and Anue), and further to the north, beyond the Belate Pass, a great basin which actually consists of two smaller basins: in the west, Bortziriak-Cinco Villas (Eng. 'Five Towns') —along which the Bidasoa River flows—, and, in the east, the Baztan Valley. To the northwest of Bortziriak-Cinco Villas, we find the north-easternmost end of (the administrative territory of) Guipuscoa, which was also included within North.-HN by Bonaparte.

In his definitive classification of Basque dialects, then, Bonaparte divided North.-HN into 6 subdialects: 1) Guipuscoan Navarrese (with the only variety of Irun); 2) Bortziriak-Cinco Villas (with the only variety of Bera); 3) Baztan (with the only variety of Elizondo); 4) Araitz (with the only variety of Intza); 5) Arakil (with the only variety of Uharte-Arakil); and 6) Ultzama (with the only variety of Lizaso). Araitz is to the west of the Ultzama Valley, between this area and Guipuscoa. Arakil is the aforementioned second stretch of the corridor (also called Middle Sakana).

In 1972 (apud *EDH*: 153), the whole of the North.-HN. territory had 104,440 inhabitants, and 51,012 —approximately half of them— were Basque-speaking. Nowadays, the vitality of Basque in this territory varies according to the area. While in Bortziriak-Cinco Villas —oriented towards the Guipuscoan Beterri— and Baztan —oriented towards Labourd— Basque is strong (obviously depending on the size of the town/village, etc.), in Doneztebe (a southern town in the basin to the north of the Belate Pass) the situation is already worse (Zelaieta 2008: 36), and in the subdialect of Ultzama (south of Belate) it is worse still. The sociolinguistic state of this subdialect, as described by Ibarra (1995: 585-595), is quite representative of what has happened and is happening with Basque in many regions of Navarre (and outside Navarre). Sixty or seventy years ago, agriculture and cattle farming was the mainstay of the economy in those valleys. This favored, in spite of the prohibition of Basque in public life, the survival of Basque, as each family constituted a basic unit, with secondary branches of the family often living in the same house. Furthermore, economic life moved around a close circle, as economic relations mostly took place with

(1998), and the language behind it appears to be, like the Basque of the Burunda region, a transition between the Basque of the west and the Basque of Navarre. Cf. now Aldai (2011b) and Ulibarri (2011) for further discussion on the Basque of this area and Port.'s language.

neighboring (Basque-speaking) villages and towns by means of fairs, markets, etc. In the last few decades, by contrast, this economic sector has lost ground and others, such as industry and tourism, have burst in. With Pamplona-Iruñea (with a population of ca. 250,000 inhabitants in 1995) being so near (to ca. 20 kilometers), the urban influence is inevitably noticeable, and many people from the capital have built a second home in the valley. The birth rate has dropped, and as a result of all these factors many small villages have depopulated.

Nowadays, schooling in Basque partially makes up for this steady decline of the language, but as Ibarra (1995: 587 ff.) points out, it is unavoidable to find the well-known symptoms of language death throughout the Ultzama Valley and its surroundings: deterioration of the language among young speakers, who often understand but have difficulties in actively speaking Basque, reduction of the lexical richness, and some other phenomena which handbooks relate not only to language death but also to pidginization and similar phenomena. Moreover, even when schooling succeeds, the Basque that comes out of it suffers from extreme code-switching, and entails a great reduction of the stylistic richness typical of the autochthonous dialect (in this case North.-HN). A case in point is the *hika* treatment (§9.6), the use of which is commonly abandoned by most speakers whose mother-tongue is not Basque. As a result, the *hika* treatment has been drastically reduced in this subdialect.

A phonetic feature characteristic of at least one region of North.-HN is the strong accent on the second syllable (from the left), in non-monosyllabic words which are not accentually marked. This corresponds to accentual Type IV in Michelena's (1987 [1972c]) classical classification, which is typical of Bortziriak-Cinco Villas and the North.-HN. Guipuscoa, but also reaches some southern regions. This applies even to Spanish paroxytonic loan words (*primábera* 'spring' < Sp. *primavéra*) and proper names (*Akílino* < Sp. *Aquilíno*), whose accent shifts to the second syllable. In this accentual type, there is a great difference in intensity between accented and unaccented syllables, which provokes frequent loss of post-tonic syllables: *amártatik* 'from 10' < *hamarr-etatik*, *garízma* 'Lent' < *garizuma*, etc. This tendency reaches its peak in Arantza (southwestern Bortziriak-Cinco Villas), where we can find syncopes which go beyond the scope of the word, often along with apheresis: *kustnu* '(s)he sees' < *ikusten du* (Zelaieta 2008: 126). There are reasons to believe that this accentual type was more widespread in the past, since in Urt.'s manuscripts (in the last quarter of the 17th century) a similar accent is assigned to the Labourdin coast (apud Michelena 1987 [1972c]: 241). In fact, although this North.-HN is a central dialect, and hence in principle innovating, this accentual type appears to be archaic, and it could be a remnant of the accent on the second syllable postulated by Michelena (1988 [1957/58]).

Another feature which is associated with North.-HN is the presence of demonstratives in which the case marker is added to the abs. form in the plural and not to the demonstrative root, such as *ok-en* 'of these' (SB. *hau-en*), *ok-etako* 'of these' (SB. *hau-etako*), *ok-etan* 'in these' (SB. *hau-etan*), etc. In Ultzama, these analogical forms only reach 1st grade demonstratives (Ibarra 1995: 328), whereas in the Baztan Valley they also reach the 3rd grade (Salaburu & Lakar 2005: 68-69). Forms like these are already attested in Old L in the 17th century, as in Ax., but they must be an innovation (§7.2.3).

The transition from G to North.-HN coincides with the transition of the Nor-Nork auxiliary *det* to *dut*, although in some towns and villages both can be used. The tripartite division B. *dot* / G. *det* / east. *dut* (S. *düt*) entails two isoglosses, and these appear to converge in the Sakana. In Urdiain (Burunda), the most frequently employed forms are *do-t* and *dau* when the subject is 1st and 3rd p., as in B, but if the subject is 2nd p. we have *de-zu*, as in G (Zuazo 2010b: 134). In Etxarri-Aranatz (Middle Sakana), there is 1st p. *de-t*, 2nd p. *de-k/n*, as in G, but 3rd p. *dau*, as in B. In Ergoiena (a valley that sticks into the Mountain Ranges of Urbasa and Andia, branching off the corridor to the south), we find the west. forms *do-t*, *do-k/n*, *do* (< *dau*), but in the towns of the Middle Sakana we already find the HN. forms: *du-t*, *du-k/n*, *du* (Erdozia 2001: 260).

Also within the finite verbal morphology, in the North.-HN. territory we have the spread of the abs. pluralizer *-it-* from the Nor-Nork to the Nor-Nori-Nork, so forms like *n-it-ion* 'I had to him/her (several things)' (SB. *ni-zki-on*) are found. This feature also appears in the South.-HN. territory, with the exception of the attestations from Gesalatz (Camino 2004: 88). Another feature shared by North.-HN and South.-HN is that of the Nor auxiliaries of 2nd p. sg. (informal) beginning with *y-*, such as *yaiz* 'you (sg.) are' (SB. *haiz*), *yitzen* 'you (sg.) were' (SB. *hintzen*) (cf. Ibarra 1995: 416-417, for the subdialect of Ultzama), or in other finite forms with an initial morpheme corresponding to that person (ibid.: 435, 437, 480, etc.). Both of these features are clear innovations.

2.2.2.5. Southern High Navarrese

Southern High Navarrese is the dialect, as always according to Bonaparte's fourth classification,[16] which was spoken in the plain around Pamplona-Iruñea

16 In fact, some other authors such as Michelena (1987 [1972c]: 238) and Ibarra (1995: 693) have proposed that, were the linguistic diversity present in Navarre to be reduced to a single binary opposition, the classification west vs. east would be more accurate. The latter has proposed a concrete classification in which an Eastern HN would comprehend Bonaparte's entire South.-HN, Baztan, Ultzama and Bortziriak-Cinco Villas. Western HN, in turn, would comprehend Arakil, Basaburua (both of them), Imotz and Araitz.

and some surrounding valleys. He further split the dialect into three subdialects: 1) western South.-HN (to the west of Pamplona-Iruñea, comprehending the varieties of Gulina, Oltza and Zizur); 2) eastern South.-HN (to the east, comprehending the valleys of Egues, Olaibar, Erro and Artzibar, and the town of Auritz-Burguete, hence 5 varieties); and 3) Izarbeibar (some kilometers to the south of Pamplona-Iruñea, comprehending the variety of Gares).

This dialect can be virtually considered to be disappeared. In 1972 (apud *EDH*: 158), the whole of the South.-HN. territory had 51,378 inhabitants (not including Pamplona-Iruñea), and only 592 of them were Basque-speaking. When Prince Bonaparte visited Navarre, however, South.-HN was still a vigorous dialect, albeit with clear signs of decline, as Campión confessed in 1884 (apud Pagola 1991: 164). Since then, the drop in the number of speakers of Basque has been vertiginous. To give just one example, in Gulina the proportion of inhabitants / speakers of Basque was 563/510 in 1860, 583/145 in 1935, and 266/20 in 1972. Some of the reasons are summarized by Camino (2004: 6 ff.). While Pamplona-Iruñea has probably been the nerve center of Navarre ever since its foundation in the 1st c. BC, until the 19th century the demographic proportion between the capital and the towns and villages throughout the whole of Navarre was not as unbalanced as it is today. Pamplona-Iruñea had much fewer inhabitants, and the villages and valleys were much more populated. In fact, most people lived bound to their land with livestock, which, unlike today, hardly left time for rest or the chance to go to Pamplona-Iruñea. Although this was the administrative center, the administrative control was not directed from the capital as much as it is today, and very occasional visits would suffice to solve any issue.

The depopulation of towns/villages and valleys, along with the coercion against Basque in public life from the Spanish central authorities in certain periods, brought about a massive loss of the South.-HN. varieties. Many people moved to Pamplona-Iruñea, but here Basque was clearly minoritary, so they abandoned Basque. This process entailed not only the disappearance of those linguistic varieties, but also of an entire conceptual world which included customs, folklore, memories of place names, etc.[17]

However, although today South.-HN is virtually a defunct dialect, we know it quite well throughout quite a long time period. Apart from direct oral records which in many cases have been made for as long as each variety has survived,

17 Even today, it is common that toponymists who gather small toponyms from depopulated Navarrese villages find out that some old person who once lived in an abandoned place is still alive and has his/her residence in Pamplona-Iruñea, so that they visit him/her in the capital to get information.

Bonaparte gathered contemporary texts, and also texts which had been written earlier (such as some by L.Elc.). Besides, many religious texts, especially sermons, have been found in churches and convents all over Navarre, most of which were written from the 18th century onwards. These, however, pose serious problems for dialectal analysis, as to find one in a given place does not necessarily imply that the priest who wrote it was native of that place, and, even if we know where he was from, we cannot know to what degree he made use of features from other varieties.[18] In addition, we have two religious works (1621, 1626) by Juan de Beriayn, native of Uterga (Izarbeibar), another one by Francisco de Eliçalde (1735), native of Gesalatz (Oltza), and a great number of sermons written between 1771 and 1821 by Joaquin Liçarraga de Elcano, native of the Egues Valley. All this can be complemented by toponymic data, which reflect few grammatical aspects, but which in turn are usually reliable as data belonging to the area where they are found.

A typical South.-HN. feature is the generalization of a columnar accent, at least in L.Elc. Thus, the position of the accent in the abs. indef. determines its position throughout the whole paradigm. Unless the word is accentually marked, that word is paroxytonic if it ends in a vowel (*zéru* 'heaven'), and oxytonic if it ends in a consonant or a semivowel (*gogór* 'hard', *artzái* 'shepherd'). If the declined word has more than two syllables, however, a second accent falls upon the last syllable if it ends in a consonant or semivowel (*zéru-arén* 'of heaven') and upon the penultimate if it ends in a vowel (*zéru-aréki* 'with heaven').

In noun morphology, a typical feature of South.-HN, which actually also occurs in Aesc and North.-HN (except in the subdialects of Bortziriak-Cinco Villas and Baztan), is the instrumental *-(e)s* instead of *-(e)z*, even if the dorso-alveolar *z* is otherwise kept. This must be an innovation (see §6.4.1.4 for a possible explanation). Another feature which includes a part of South.-HN and Aesc is the existence of demonstratives beginning with *g-*. In L.Elc.'s writings, these have *g-* when they function as determiners: *pleitu-gáu* 'this lawsuit' (SB. *auzi hau*), *gau-gártan* 'that night' (SB. *gau hartan*). But they have no *g-* when they are pronouns (*ori* 'that', SB. *hori*) or adverbs derived from demonstratives (*emen* 'here', < *hau-en*, SB. *hemen*).

In the finite verbal morphology, a feature of South.-HN, again shared with Aesc, is the lack of final *-n* in past forms: *nitza* 'I was' (SB. *nintze-n*), *ze* '(s)he was' (SB. *ze-n*), *nue* 'I had' (SB. *nu-e-n*), *zue* '(s)he had' (SB. *zu-e-n*), etc. This

18 A typical case is that of a sermon written by the priest Joseph Marcotegui. He was born in Goñi (Oltza) in ca. 1740. The sermon was found in a convent in Alegría-Dulantzi (eastern Alava), but had been written to be uttered in Ollo (next to Goñi, also within Oltza). However, in the sermon there are plenty of linguistic features which we know for sure are neither from Goñi nor from Ollo (Camino 2004: 46).

must be an innovation, which probably has its cause in reanalyses of past forms with suffixes such as the complementizer *-la* or the relativizer *-n*: *zu-e-n* '(s)he had' ~ *zu-e-la* 'that (s)he had', hence → *zu-e* '(s)he had'. The same process has apparently taken place in northwestern B (§2.2.2.1), but independently.

South.-HN —and also some North.-HN. varieties— shows forms of the type *zeki-* for past Nor-Nori auxiliaries: *zekio* '(s)he was to him/her' (SB. *zitzaion*), *zekigu* '(s)he was to us' (SB. *zitzaigun*). These forms with *zeki-* are aoristic in archaic authors such as Lç., but in later texts they have specialized in the sbjv. past. How this reanalysis of the TAM system occurred with respect to the roots **edin* and **ezan* is described in §8.4.3. The interesting point about South.-HN is that the archaic maintenance of the aoristic root for the indicative past tense has been restricted to the auxiliaries of the Nor-Nori type.

Another interesting feature of finite forms in South.-HN is that the 1st p. sg. dat. agreement marker *-da* does not become *-t* in word final, as is usual in the rest of the dialects:[19] L.Elc. *atzendu zeki-da* 'I forgot it', L.Elc. *ori ez-tagoki-da ni-ri* 'that does not correspond to me'.

2.2.2.6. Labourdin

Labourdin is the westernmost dialect in the continental Basque Country. In Bonaparte's classification of 1869, it occupies the western part of the administrative territory of Labourd, and also Urdazubi and Zugarramurdi, to the north of Baztan, in Navarre. In fact, Yrizar included the whole of Baztan within Labourdin, taking into consideration the hesitations expressed by Bonaparte himself in *Le verbe basque en tableaux* and in a comment of 1881 (apud Yrizar 1991-2008: Labourdin, 2). In any case, Bonaparte's fourth classification comprehended two subdialects: 1) the proper one (with the varieties of Sara —including Urdazubi and Zugarramurdi—, Ainhoa and Donibane Lohizune); and 2) the mixed one (with the variety of Arrangoitze, to the northeast of the Labourdin linguistic territory).

Around 1972 (apud *EDH*: 272), there were 22,750 speakers of those four varieties, with 16,240 of them in the variety of Donibane Lohizune. Labourd is the most urban territory of the continental Basque Country, as the coast attracts a great deal of tourism from France and Spain. Traditionally, the Basques from the continental territories have had to resort to emigration. At the beginning of the 20th century, most of them emigrated to America (especially to Argentina and

19 Outside South.-HN, similar forms also appear e.g. in Ultzama, as in *zai-de*, *zai-re* '(s)he is to me', < *zai-da*, with vowel harmony and rhotacism in the second variant (Ibarra 1995: 422), and in Aescoa (see Camino 2004: 88-89 for the exact extent).

Uruguay), but later the point of attraction moved to large French cities like Paris or Bordeaux. Along with this outward migration, a parallel process of inner migration from east to west has taken place within the continental Basque Country, as a result of which an important population center has taken shape in the conurbation Biarritz / Angelu / Bayonne (near to the coast). As a consequence, nowadays the city of Bayonne hosts a considerable number of Basque speakers, and has become the virtual capital of the continental Basque Country, albeit having historically belonged to the Gascon linguistic domain. As for the language, the Labourdin dialect is going through a rapid process of dilution into what has come to be tagged *Navarrese-Labourdin*. The influence of Standard Basque is also breaking through among young generations.

Labourdin is particularly well documented from the 17th century, since in this century many works, especially religious ones, are written and printed, many of them in the circle which is sometimes referred to in the literature as the *School of Sara* (but cf. §3.4.2). Although Labourdin literature suffered a decline in the 18th century, the development of L can be tracked since its beginnings (Pikabea 1993). In spite of the elegance and classicism of these authors, however, the central nature of L causes it to be innovating.

A phonetic feature of L, shared with most of the LN. territory,[20] is the lack of palatalization even with *n*, *l* in the most palatalizing context: / j _ V. Thus, L. LN. *mainu* 'bath' is pronounced as [maɲ̟nu], not [maɲu], and *boteila* 'bottle' as [boteɲ̟la], not [boteʎa] (*FHV*: 195-196). In recent loan words, however, the palatalization is not avoided (Zuazo 2008: 165-166). As Pikabea (1993: 29-42) has shown, the orthography employed by 17th century L. authors reveals that the depalatalization process is an innovation in L, as they use <ñ> and <ll> not only in the context / j _ V, but also after a syllabic *i*, i.e. in the context / i _ V. Thus, Ax. writes *gaiñean* 'on the top', *irakurtzaille* 'reader', *milla* 'thousand', etc., and so do all authors of the 17th century and most of the 18th century. Generally speaking, it is in the last quarter of the 18th c. that we have to set the limit from which at that point onwards L. authors begin to reflect depalatalization in orthography, writing *gainean*, *irakurtzaile*, *mila*.[21] According to Pikabea, depalatalization has

20 Cf. Zuazo (2008: 245). There it is shown that, in the Burunda, the same depalatalization phenomenon is found.

21 Although this is generally the case, it can vary depending on the word concerned, at least in Ax., the author on which Pikabea bases his research. With the word 'but', for instance, Ax. shows hesitation between *baiña* and *baina*. Another curious phenomenon present in Ax. is that, in the context / i _ # (i.e. after syllabic *i* and word final), *n* and *l* are always written with no orthographic palatalization: <min> 'pain', <muthil> 'boy'. Even when these are declined or derived, they are written unpalatalized in most cases: <min-ak> 'pains', <muthil-ak> 'boys' (but <hill-ak> 'dead (pl.)').

spread due to the influence of Low Navarrese. In Dch., non-palatalized orthography is systematic: <hil-en> 'of the dead', <hilabete> 'month', <konpainia> 'company'. Even Lç. often writes with non-palatalized orthography (<hil-ak> 'dead (pl.)', <baina> 'but'), but not systematically (<billuzi> 'naked', <oillar> 'cock'). It is probable that LN is the most archaic dialect in this respect. If this were the case, although depalatalization also occurs in L, this dialect would not have acquired this feature directly from the common language.

Both in noun and verb morphology, if we move from west to east along Biscay, Guipuscoa, and then into the continental Basque Country (leaving Navarre aside), L marks the beginning of the eastern linguistic realm. We find erg. pl. *-ek* (west. *-ak*), dat. pl. *-ei* (west. *-ai*), which will compete with *-eri*, and even with the far east. *-er* (at least in written literature), from the 19th century onwards (Pikabea 1993: 48), all. *-ra(t)* (often with both forms alternating in the same author; §6.4.2.2), *-a(ren) baita-* for the local cases of animate nouns, etc.

In the finite verbal morphology, some Nor auxiliaries of L have a special vocalism: *zare* 'you (sg.) are' (SB. *zara*), *gare* 'we are' (SB. *gara*), *dire* 'they are' (SB. *dira*), among others. These appear from the earliest texts, albeit not exclusively (Ax. e.g. has *zara*, *gara*, *dira*). *Zare* and *gare* might be dissimilations, or forms built by analogy with *dire*, which in turn would have arisen by vowel assimilation (< *dira*). Alternatively, all these forms might be a result of a reanalysis of the corresponding suffixed forms (e.g. ← *zare-la* 'that you (sg.) are' < **zara-e-la*, as suggested by Ariztimuño; cf. §9.8.1, in footnote). In Nor-Nori-Nork auxiliaries, from Labourd to the east along the continental dialects, the root is *-erau-* or phonetic developments of this. This root comes from **-eradu-*. Apparently, the common language had a procedure for turning Nor-Nork auxiliaries in **da-du(n)* into ditransitives of the Nor-Nori-Nork taking a dat., whereby **da-du(n)* → **de-ra-du(n)*, with insertion of *-ra-* and an *-a-* > *-e-* change of the pre-radical vowel. The root *-i-*, however, has prevailed in L when the dat. verbal index after the root is of 3rd p.: *d-i-o-t* 'I have to him/her', *d-i-o-gu* 'we have to him/her', etc.[22] Inversely, although *-i-* has prevailed in most areas of Navarre, the root *-erau-* in Nor-Nori-Nork auxiliaries has prevailed in Baztan and a part of the Sakana. Besides, in the HN. authors of the 16th and 17th centuries, both roots alternate in a facultative fashion (Kerejeta 1991: 168), and we have already seen that we also have Nor-Nori-Nork auxiliaries with the root *-erau-* in Old G (with abs. pl.) and Al (§2.2.2.1).

22 In Ax. and some other early authors, however, there are some Nor-Nori-Nork auxiliaries with *-erau-* with 3rd p. dat. indexes, so that forms like *derauka* and *dio* are apparently in free variation.

A recent innovation, but still characteristic of L —albeit not unparalleled in some surrounding areas, and also in Lekeitio (B)—, is the (not systematic) use of Nor-Nork auxiliaries for semantically tripersonal clauses, as in *erran nau* '(s)he has said it to me' (vs. SB. *erran dit*).

Let us conclude by saying that, in spite of the innovating nature of L, at least in the maintenance of the Linschmann-Aresti Law, Old L appears to be archaic, as Ax. and E.Cib. fulfill the law systematically, and unlike LN. and S. authors they have not lost the strong pronouns *zeure*, *geure*, *zeuen* (§7.3.4).

2.2.2.7. Western Low Navarrese

According to Bonaparte's fourth classification, Western Low Navarrese comprehends a part of east. Labourd, the southwest. quarter of the Low Navarre, and the Aescoa Valley in the peninsular Navarre. These correspond to the following subdialects: 1) West.-LN from Labourd (with the varieties of Mendionde-Lekorne and Uztaritze); 2) Baigorri (one variety: Baigorri); and 3) Aescoan (one variety: Aescoa). Around 1972, in this territory there were 22,793 speakers of Basque, out of a total number of 30,230 inhabitants (apud *EDH*: 321-322).

The division of Low Navarrese into two dialects, both with a status of such, has been controversial. Azkue merged them into one, and Bonaparte himself admitted that they could equally be considered as a single dialect (apud Yrizar 1991-2008: West.-LN, 2). Furthermore, Low Navarrese is generally poorly documented between 1545 and 1782 (see Camino 2008: 14), and hence it is difficult to discern how widespread the features which appear in Dch.'s work of 1545 could be throughout the whole of the Low Navarre.

In any case, an important set of isoglosses which delimits L from West.-LN and all the regions to the east of the continent concerns the vocalism of some finite verbal forms. In Nor auxiliaries, while L has *naiz* 'I am', to the east we have *niz*, a monophthongization of a form of high frequency in discourse. Similarly, while L has *zare* and *gare*, to the east we have *zira* and *gira*. Rather than a dissimilation like in G. *zera*, *gera*, in this case we appear to have an analogical process under the influence of *dira* of the same paradigm. Some pot.-sbjv.-imp. forms of the Nor type, built upon the root **edin*, are also monophthongized from West.-LN to the east: *zite(ze)n* 'you (sg.) are (sbjv.)' (L. *zaite(ze)n*, SB. *zaitezen*), *gite(ze)n* 'we are (sbjv.)' (L. *gaite(ze)n*, SB. *gaitezen*), and so are some forms of the Nor-Nork type: *nu* '(s)he has me' (L. SB. *nau*), *zitut* 'I have you' (L. SB. *zaitut*), *gitu* '(s)he has us' (L. SB. *gaitu*).[23]

23 I have taken these forms from the grammatical appendices to Bonaparte's manuscripts made by Pagola *et alii* (eds.; 1992-1999), corresponding to L and West.-LN. However,

We also have L. *natzai-* vs. *nitzai-* to the east in Nor-Nori auxiliaries, and L. *dio* vs. *dako* (< **derauko*) to the east (but *deio* in S) in Nor-Nori-Nork auxiliaries.[24] In some areas like Uztaritze, Zugarramurdi and Baztan, *dio* and *dako* are still competing (Zuazo 2008: 20).

In the peninsula, Aescoa, along with Salazar and Roncal, forms a valley belt in the northeast of Navarre, the (today largely disappeared) Basque language and ethnological culture of which are of the utmost interest, due to their archaic nature. As in Salazar and Roncal, the economy of Aescoa is or was primarily based on pasturage (Camino 1997: 84-91). The need to protect the grazing lands of the valley from shepherds from other valleys creates a strong sense of belonging among its inhabitants, which makes the valley a political unit in itself, often with administrative institutions and sometimes even with its own army. This often results in the use of micro-ethnonyms for the populations of each valley. Thus, while the inhabitants of Aescoa are referred to as *aetzak* —from which the name of the valley derives—, they call those from the neighboring Urraulgoiti and Artzibar *pardixak* ('partridges').

As a dialect, Aesc has long been declining. Twenty years ago, it had 342 speakers out of 1,115 inhabitants (ibid.: 111-137). Linguistically, it is a crossroad of all the surrounding varieties (ibid.: 153-261). Bonaparte's inclusion of the dialect within West.-LN was due to his emphasis on finite verbal morphology as a delimiting criterion. Since Aesc has forms like *niz* 'I am', *zira* 'you (sg.) are' and *gira* 'we are', it was related to LN, but, in fact, within this same realm, Aesc also has some important forms which align with the peninsular varieties, and move away from the corresponding LN. ones, such as LN. *nintzan* 'I was' vs. Aesc. *nintze*, HN. *nitze*, and LN. *zako* '(s)he is to him/her' vs. Aesc. *zeo*, HN. *zaio* (ibid.: 246). We have already seen that, regarding some other important features, Aesc is also peninsular. With South.-HN, in particular, it shares the instr. *-(e)s*, the lack of the *-n* of the past in finite verbal forms (although there is often free variation between both variants, ibid.: 247), and the demonstratives beginning with *g-* (not with the whole South.-HN. territory; cf. §2.2.2.5). In these, however, the distribution between the variants with and without a stop (*gura* ~ *ura*, etc.) is very difficult to predict, unlike in L.Elc., and although some tendencies can be drawn, they appear to have become free variants (ibid.: 410-414).

---

each form must be treated individually, since, as Pikabea (1993: 57-60) notes, the logical correspondence L. *daite(ke)* '(s)he can be' / LN. *dite(ke)* does not work. For example, Ax. always has *diteke*, and Pouv. most of the time (cf. *CA*).

24 Some Labourdin authors, however, have employed LN. forms since the 19th century (Pikabea 1993: 50-66).

2.2.2.8. Eastern Low Navarrese

Eastern Low Navarrese, in turn, comprehends a part of Labourd, the east and north of the Low Navarre and the Salazar Valley in the peninsula. These correspond to the following subdialects: 1) East.-LN of the Adour (with the varieties of Beskoitze and Urketa); 2) Cizain-Mixain (with the varieties of Cize, Mixe, Bardoze and Arberoa); and 3) Salazarese (with only one variety: Salazar). Around 1972, in this territory there were 23,782 speakers of Basque, out of a total number of 31,764 inhabitants (apud *EDH*: 323-325).

As I have said, the need to split LN into West.-LN and East.-LN, both with the status of dialect, has been questioned. The differentiating factor which caused Bonaparte to create an East.-LN was the existence, from this region to the east (i.e., in East.-LN and S, including Sal and R), of allocutivity, also with the *zuketa* treatment (§9.6.3). Thus, when a speaker is using this treatment, (s)he has to insert an index of the hearer *zu* into the finite verb even if this is not a core argument of the sentence, just as in the *hiketa* treatment of all of the dialects (*zu zira* 'you (sg.) are', but *hura da* '(s)he is' → *hura du-zu*). This, however, is not systematic in Dch., and Lafon (1999 [1951]: 754) regards this lack of regularity as evidence that it is an innovation which was developing at that time. In addition, a feature exclusive of East.-LN (including Sal) is the creation of a third type of allocutive treatment, the formality level of which is intermediate between *hika* and *zuka*, which consists in the meaningful palatalization of *zu* and its verbal indexes (*xu xira*, *hura du-xu*).

From the point of view of a dialectal analysis which focuses on what the earliest stages were like, the crucial source for East.-LN is Dch. (cf. Lafon 1999 [1951] for a study of his language), because this author was from near Donibane Garazi, he presents a corpus of an acceptable size, and LN is not attested again until much later (leaving aside what Lç. may have of LN). The vocalism of finite forms that we have seen for the West.-LN already appears in Dch. Some other features which are present in Dch. (whether they exist in West.-LN or not) and from his region to the east (i.e. in S and R) are the dat. pl. *-er* (S. *-ér*), the non-contraction of strong 1st and 2nd personal pronouns such as *ni-haur* 'I myself (abs.)', *ni-haur-k* 'I myself (erg.)', etc., except in gen. forms (§7.3.3), and the (not systematic) lack of dat. agreement in the finite verb (Dch. *andri-ari emaiten du ogen-a* 'he gives/does harm to the woman'; SB. *dio*). This feature has tended to spread westwards in later periods, through the continental Basque Country.

In its finite verbal morphology, Dch. is more archaic, at least in some aspects, than Lç. and the L. authors of the 17th century. The modal morpheme *-te*, present in Nor auxiliary verbs built upon *izan* or **edin*, has not been recharacter-

ized by the more productive *-ke* (Dch. *da-te* '(s)he can be, will be' vs. Lç. *da-te(ke)*, Ax. *da-teke*). Similarly, the abs. pluralizer *-z-* has not reached Nor auxiliaries built upon **edin* like Dch. *gite-n* 'we are (sbjv.)' (vs. Lç. *gaite-ze-n*, Ax. *gaite-zi-n*). In Nor-Nori auxiliaries, we find two forms with the root *zau-* in Dch. (albeit among many more with the more widespread *zai-*), which is LN, as opposed to L. and S. *zai-*. Perhaps it was spreading at that time. By the 19th century, it is also used by some L. authors, perhaps favored by the vocalism *dau-* of Nor-Nori-Nork auxiliaries (Pikabea 1993: 55-56).[25]

A process which conditions the noun and verb morphologies of eastern continental Basque, and which in Dch. appears to be in its earliest phase, is the set of morpho-phonological phenomena which occur when a root, whether nominal or verbal, ends in *-u-* and a flectional *-a/e-* is attached. In Dch., a yod is inserted. Thus, *mundu-* + ines. sg. *-an* > *munduyan* 'in the world', *zeru-* 'sky' + ines. pl. *-etan* > *zeruyetan* 'in the skies', *du-* '(s)he has' + rlv. *-en* > *duyen* 'that (s)he has', etc. The next step entailed the disappearance of *-u-* and the conversion of *-i-* into syllabic, forming a hiatus with the suffix: *mundian*, *zeri-etan*, *dien*. This process occurs in Cize today (or at least in Lafon's times), albeit usually with monophthongization of the hiatus (Lafon 1999 [1951]: 738). In S. authors of the 17th century, the process is already accomplished, and forms vocalic alternations within noun and verb paradigms such as: *mündü bat* 'a world' but *mündi-an* 'in the world'; *zelü bat* 'a sky' but *zeli-etan* 'in the skies'; *dü-t* 'I have' but *di-e* 'they have' (< **du-e*, SB. *du-te*); or even *dü-zü* 'you (sg.) have' but *dü-zie* 'you (pl.) have' (< **du-zue*), in which the morpho-phonological process has applied not to a root-suffix juncture but to an index-pluralizer juncture. At least in today's S and Cizain, two further morpho-phonological processes which are systematically operative are *-e-a/e* > *-ia/e* (*álthe-* + *-a* > *althía* 'the side') and *-o-a/e* > *-u-a/e* (*sórho-* + *-a* > *sorhúa* 'the meadow', actually pronounced *sohúa* in S). These are already present in Dch. but are not systematic, as can be seen in *Iangoiko-* + *-a* > Dch. *Iangoikoa* ~ *Iangoikua* '(the) God' (ibid.: 738). The fact that *-u-a/e* > *-uya/e* is systematic but *-o-a/e* > *-ua/e* is still not in Dch. indicates that, unlike the morpho-phonological processes of the west (of the push-chain type, cf. §2.2.2.1), in the east we find ourselves confronted with a drag-chain process.

On the peninsular side of the East.-LN. territory as delimited by Bonaparte, between the valleys of Aescoa and Roncal, we have the Sal. dialect. Around

25 Another interesting feature is the abs. pluralizer *-de* of the verb *izan* 'to be', as in Dch. *zira-de* 'you (sg.) are', *gira-de* 'we are', *dira-de* 'they are' (along with *zira*, *gira*, *dira*), and not because it is characteristic of that dialect or region (forms with *-de* stretch up to Al and G), but because, as Camino (2008: 104) has noted, this *-de* mysteriously disappears from later LN. tradition (again leaving Lç. aside), even though it is present in S., Sal., and R. texts of the following centuries.

1972, it had 327 speakers out of a total number of 2,173 inhabitants (apud *EDH*: 323), but today it is virtually defunct. Three of its last speakers were interviewed by Aitor Arana some years ago (at least one them in 2000), and using that information and some other sources he published a learning method of Salazarese (Arana 2006), along with a dictionary.

The main factor for Sal to be classified as a subdialect of East.-LN was, as is expected from Bonaparte, the *xuka* allocutive treatment in the verbal morphology. Whilst this decision may be arbitrary, the fact that Sal shares this feature with varieties from beyond the Pyrenees shows once more that these have never been a linguistic barrier. On the contrary, they have been linguistically and culturally permeable, due to the markets, fairs and *facerías* or treaties between shepherds from both sides.

A feature which Sal shares with R is the existence of demonstratives, and adverbs derived from those demonstratives, beginning with *k-* (§2.2.2.9). A feature exclusive to Sal is the morpho-phonological process whereby, when the article *-a* is attached to an *a*-ending root, an *-r-* is inserted (only in the abs.): *alaba* 'daughter' + *-a* > *alabara* 'the daughter'. Thus, a morphological distinction between the pure root and the abs. sg. form arises, as in S, R and the west. The other distinctive feature of Sal is the huge number of syncopes that it undergoes. This process has parallels in neighboring varieties, but they are nowhere as strong as in Sal. Examples are: *abrats* 'rich' (SB. *aberats*), *bedratzi* 'nine' (SB. *bederatzi*), *pezta* 'peseta' (SB. *peseta*). They also occur in finite verbal forms such as *gra* 'we are' (SB. *gara*, East.-LN. *gira*), *zra* 'you (sg.) are' (SB. *zara*, East.-LN. *zira*), *dra* 'they are' (SB. *dira*), *ztu* '(s)he has you (sg.)' (SB. *zaitu*, East.-LN. *zitu*), etc. As Hualde (1997b: 98) shows, Sal has moved from a paroxytonic accent, as in S, to a columnar one, as in South.-HN, so that words like *arrí-a* 'the stone' and *lekú-a* 'the place' have developed into *árrya* and *lékwa*, with transformation of the stem-final vowel into non-syllabic (also of *-e*, *-o*). In the noun declension, the syncope must have preceded the accent shift. Hence, the evolution must have been something like: *korralé-a* 'the barnyard' > *korlé-a* (syncope) > *kórlĕa* (accent shift), or *apostorú-a* 'the apostol' > *apostrú-a* (syncope) > *apóstrwa* (accent shift).

2.2.2.9. Souletin

Souletin is, perhaps after B, the most idiosyncratic Basque dialect. It is quite well documented from the 17th century onwards, thanks to authors like Tt. and Oih. In Bonaparte's fourth classification, it was divided into two subdialects: 1) the proper one (with only the variety of Atharratze, which includes almost the whole territory of Soule); and 2) Roncalese (with the varieties of Bidankoze,

Urzainki and Uztarroze, in the peninsular Basque Country). In Soule, in around 1972, the number of speakers of Basque was 11,095, out of a total number of 18,839 inhabitants (apud *EDH*: 438). At least until recently, there have been some neighborhoods and a village, Esquiule, to the northeast of but outside of Soule, in Bearn, where Basque was/is partially spoken. In the last decades, however, S has been a dialect in steady decline, and the number of speakers has considerably dropped. Depopulation has been huge due to the crisis of the primary sector, and French central authorities have shown no interest in fostering it at school (see Coyos 2007: 53-62, for more details about the sociolinguistic situation).

Soule has a high region, next to the Pyrenees, called Basabürüa, in which the main villages are Larraine (in the southwest) and Santa Grazi (in the southeast). The low region is flatter and has more towns, the capital Maule being towards the north. Souletins are known for having quite a strong awareness of themselves as an idiosyncratic territory. They call their neighbors from Cize *manexak* (*manex* < *(Done) Ibanes*, 'Saint John' in Basque, as this is the patron saint of Donibane Garazi), and, by extension, all continental Basques from outside Soule.

Roncal is the valley on the northeast. corner of Navarre. In its northern part (variety of Uztarroze), it has the towns of Uztarroze and Izaba. In a central part (variety of Urzainki) there are two more towns: Urzainki and Roncal; and in the southern part (variety of Bidankoze) three more: Bidankoze (to the west), Garde (to the east) and Burgi (at the southern extreme). When Bonaparte visited the Roncal Valley in 1866 (López-Mugartza 2008: 159-174), R was partially spoken in all of those towns, but in those of Bidankoze it was about to disappear. In Bidankoze itself, the last speaker —Mariano Mendigacha, who had been Azkue's informant— died in 1931. Yet even in the 1950s, Michelena and some collaborators visited the valley with a large tape recorder, and had the opportunity to carry out a linguistic interview with an old man and two old women in Uztarroze and three old women in Izaba (the material was published in Michelena 1988 [1953b]). A few more testimonies were gathered in the following years, but by 1980 the dialect had died out... Or it was believed to have died out, as some years later it was discovered that a Roncalese-speaking old woman from Uztarroze, Fidela Bernat, still lived in Pamplona-Iruñea. Ricardo Gómez spoke with her in 1988, and in 1990 she was interviewed by a local radio station. She died in 1992, and with her the last glowing ember of the R. dialect went out for ever.[26]

26 Note that, in general, the last speakers of R were women. The traditional explanation for this is that, until not too long ago, in winter, while many men from the Roncal Valley

Both the S of Soule and R are linguistically very important for diachronists due to a number of archaisms. Most of these archaisms, however, are phonetic. One of them, not present in R, is the maintenance of the aspiration (of /h/, but also as a feature of /p^{h}/, /t^{h}/, /k^{h}/, and in the clusters *-nh-*, *-lh-*, *-rh-*), which nowadays in Soule is the clearest of all the continental dialects. This can only occur once per word, and never to the right of the second syllable (from the beginning). It was closely related to the accent, at least until prehistoric S (§4.6.3, §4.6.4). In fact, the accent is another archaic feature. Apart from being quite strong, there are clear traces that, some marked words left aside, the S. accent was once paroxytonic (always on the penultimate syllable), and is still so in most of the lexicon. Some contractions, however, have created some oxytonic words, and in some cases morphological alternations. For instance, while the abs. indef. of *alhába* 'daughter' is still the same, the abs. sg., with article, was once **alhabá-a*, but has contracted into *alhabá* 'the daughter'. Thus, a morphological distinction —which also exists in R— between the pure root and the root plus the article has arisen, just as in Sal and in the west (see Zuazo 2008: 49-51 for more alternations).

A feature shared by the S of Soule and R is the maintenance of the voiceless stops *k*, *t* in the context / n, l _, as in *igánte* 'Sunday', *ált(h)e* 'side', etc. (vs. *igande*, *alde*, in the other dialects). This is traditionally considered as an archaism, but Michelena (*FHV*: 355) proposed another possibility (cf. §2.3.5). An evident archaism of both dialects is the maintenance of nasalized vowels, although this feature is disappearing among young Souletins. This is extremely interesting for diachronic research, as it appears that the intervocalic loss of *-VnV-*, almost common to the whole territory, > *-Ṽh̃Ṽ-* > *-Ṽ̃Ṽ-* > *-VV-*, has not reached the last stage on this geographic corner.[27] Thus, for instance, the fact that R maintained nasalized vowels in words such as *õl* 'plank' (vs. *o(h)ol* in other dialects) and *ãr* 'worm' (vs. *(h)ar* in other dialects, Old B. *haar*) tells us that there was once a *-VnV-* sequence there: **onol*, **anar*, as reconstructed by Michelena.

---

were shepherds who moved as far south as the Bárdenas Reales (south. Navarre) to new pastures for their cattle, and consequently became speakers of Romance, many women typically moved to Maule (Soule) to work in espadrille factories (Basq. *espartin*), so they continued speaking Basque. These young women were called *enarak* 'swallows', because their migration cycle recalled that of swallows, and because they wore a black dress.

27 The penultimate stage of this development is paralleled in Old B, for which we have, among others, the evidence of graphic notations such as Gar. *ardão* 'wine' (< **ardano*), Cap. *erraçoẽa* 'the reason' (< Lat. *ratione(m)* + *-a*), etc. (cf. *FHV*: 48-49, 301).

An innovation of S (including R) is the change *-au-* > *-ai-*, except before *r*, *rr*, *s*, *ts*, and after *j-*. Thus, we have S. R. *gaiza* 'thing' (vs. SB. *gauza*), *gai* 'night' (vs. SB. *gau*), but *gaur* 'today; tonight', *jaun* 'sir'. An innovation of S, also present in Mixe, which does not include R, is the change *u* > *ü* except before *r*, *rd*, *rth*, *s*. Thus, S. *negü* 'winter', but *úrthe* 'year'. In turn, this change has provoked the subsequent *o* > *u*, but this change is widespread in other continental dialects (although not as strongly as in S),[28] and its phonetic contexts are not so easy to define (Zuazo 2008: 44-45).

A peculiar feature of R, not present in the S of Soule, is the presence of *k-* in demonstratives (1st grade *kaur*, 2nd grade *kori*, 3rd grade *kura*, etc.), or adverbs which come from demonstrative roots (*keben* 'here', < **kau-en* '~ in this'). We have already seen this feature in Sal (§2.2.2.8), and *g-* in South.-HN (§2.2.2.5) and in Aesc (§2.2.2.7). This is widely believed to be an archaism, inasmuch as the change k^h-, t^h- > *h-* did not reach these forms. Note, however, that outside these demonstratives (or adverbs etymologically derivable from demonstratives) there is no further correspondence R. Sal. *k-* / Aesc. South.-HN. *g-* / *h-* ~ *Ø-* in the rest. The archaism, if there is such, must have been morpho-syntactic rather than phonetic, as otherwise we would expect to find more alternations. In other words, a form like *kori* 'that' may have survived because in R and Sal the sequence *uso-kori* 'that dove', with *-kori* as a determiner, may have been interpreted as an accentual unit, and hence *-k-* was not in word-initial position. In other dialects, *uso kori* was interpreted as two words, and hence the change *k-* > *h-* was accomplished.[29] The forms with *g-* of Aesc and South.-HN are perhaps due to the regular initial voicing of stops (cf. §4.4.2.2), after a later reinterpretation of *uso-kori* as two words.[30]

---

28 *ü* itself is not unknown in the other continental dialects, but it appears only in loan words from French.

29 Alternatively, we can consider that in the other dialects the demonstratives in pronominal function, in which the change occurred (*kori* > *hori*), have been generalized and replaced the conservative determiners with stop.

30 This is approximately the standard view. However, as Lakarra (p.c.) reminds me, the varieties which show word-initial stop in such cases begin to be attested quite late. In fact, the letters from 1615-17 written in S and R which are currently being edited (cf. §3.4.2) do not show such stops in the expected environments. An alternative approach, then, would be that the variants with stop are innovations. In such a case, they could be similar to alternances present in suffixes like *gos-ari* 'breakfast' vs. *ba(ra)z-k-ari* 'morning snack', *aha-ide* 'relative' vs. *auzo-k-ide* 'neighbor'. As loan words show (*gorr-eria* 'deafness' vs. *tontu-k-eria* 'silly thing', < Sp. *-ería*), the stopless form of the suffix must be older, and hence perhaps the suffix with stop can be explained as the result of an epenthetic process.

In the realm of morphology, by contrast, S and R are not so peculiar. Or to put it differently, the particular nature of some morphological elements are the result of the appearance that they have acquired as a result of some phonetic changes. This is the case of the vocalism of the Nor-Nork auxiliaries in S (*dü-t* 'I have', etc.), of the 1st p. sg. erg. agreement marker of R (*du-d* ~ *du-r* < **du-da*), and of the vocalism of some Nor-Nork auxiliaries in S which are the result of the change *-au-* > *-ai-*: *nái* '(s)he has me' (< *nau*), *nái-e* 'they have me' (< *nau-e*; SB. *nau-te*), etc. A morphologically peculiar form —or rather set of allomorphs— is the morpheme *-ala(t)* and variants, present in S and Sal, which is peculiar because of the presence of the article in a local case and of a lateral instead of the general rhotic (R presents an allomorph *-ara* with article, but none with *-l-*). The interpretation is complicated, but it seems that we have to start from a morphological form different from the *-ra(t)* of other dialects (§6.4.2.2).

2.2.2.10. Some other features: A-B-B-B-A and A-A-A/B-B-B alignments

In this subsection, I intend to present some other features which have remained unmentioned in the one-by-one analysis of dialects because they do not belong to just one dialect or area, but can occur in areas located at opposite extremes, or in one half of the Basque-speaking territory, or can form a *continuum* along the whole territory. The diachronist can divide these alignments into two types (which are in any case approximate): 1) alignments of the type A-B-B-B-A, in which a feature A is kept at two opposite ends of the territory, and is therefore archaic, whereas the central territory has innovated creating B; or 2) alignments of the type A-A-A/B-B-B, in which one part of the territory has the feature A, and the opposite part the feature B, with a region in which both features are possible.[31]

Among A-B-B-B-A alignments, very typical are some of the features which turn up in both Biscayan and Souletin, but not in central regions. An example is the archaic inessive *-n*, where the article *-a-* has not spread, as it would originally have corresponded to a local case (B. *itaurre-n* 'leading oxen', S. *etxe-n* 'at home'; cf. §6.4.2.1). The conservative regions at opposite ends, however, are not reduced to these dialects, and their extension varies depending on the feature. In the verbal morphology, we have the past forms beginning with *l-* in subordinate clauses corresponding to main clauses with the verb in a past or irrealis tense (Lafon 1999 [1956b], Aldai 2008; cf. §9.4.2), as in Dch. *[Gaiz-ez lago-la] enzun nuien* 'I heard that he was angry'. Among authors with an extensive corpus,

31 Of course, there are more conceivable types of alignment, but I think these ones are useful and sufficient for summarizing the most relevant remaining features.

these appear in Dch. and Lç. (in the east), and now in Laz. (in the west), but they later disappear, turning to the *consecutio temporum* present in neighboring Romance (as in SB: *[gaitz-ez zego-ela] entzun nuen*). Another archaic feature which appears in early and mostly peripheral texts is the aoristic (but indicative) meaning of the auxiliaries with the roots **gin*, **(e)zan* (transitive) and **din* (intransitive), which after the reorganization of the TAM system became restricted to pot., sbjv. and imp. forms (cf. §8.4.3).

Another feature of the verb morphology which is present at both ends of the territory is the 3rd p. pl. erg. agreement marker, which is *-e* in western and eastern dialects, but *-te* in central ones (B. *dab-e(-e)*, S. *di-e* 'they have it', vs. North.-HN. SB. *du-te*). Both variants must come, according to Trask's (*HB*: 221) widely accepted proposal, from *-de*, attested in Old B. *dau-de* 'id.'.[32] From devoicing contexts (after a sibilant), *-te* —originally an allomorph— was generalized, and from intervocalic contexts after the dropping of *-d-*, the allomorph *-e* was generalized. According to its geographic distribution, the latter generalization was probably earlier. We can also mention here the binding vowel which links the end of the verb to the *-n* (relativizer or of past) and *-la* (complementizer, etc.). In some finite forms, that vowel is *-a* in the west and the east, but *-e-* in central regions (B. *neban*, S. *nian* 'I had it', both < **ne-dun-a-n*, vs. North.-HN. SB. *nuen* < **ne-dun-e-n*) (cf. §9.8.1).

A syntactic feature with A-B-B-B-A alignment is the nature of the NP which triggers abs. pl. agreement on the verb.[33] In old western texts and in current far eastern varieties, that NP must be countable, mathematically plural (two or more), and definite (RS. *Zaran bat dagi-a-nak bi dai* '(s)he who makes a basket, will make two', Tt. *deabru-ak ba-di-zu bi gauza* 'the devil has two things'). This feature must be old, while the extension of the abs. pl. agreement to more contexts (e.g. to countable, mathematically plural, but indefinite NPs) must be dialectal (*HB*: 221).

In alignments of the type A-A-A/B-B-B, by contrast, the diachronist must consider with care which side of the territory has been conservative and which innovating, although it can also happen that A and B are different elections from a common pool of possibilities present in the common language. In the noun

32 Actually this form or similar ones still survive in some places, as *dau-de* in Ugao-Miraballes (Salazar 2001: 67), the pleonastic *dau-dee* (and *dau-die*, *dau-dei*) in Orozko and its surroundings (Ulibarri 2008: 201), or *eu-de-n* (SB. *zu-te-n* 'they had') in Gatika (Gaminde 1997: 87).

33 Perhaps the issue is similar with any core argument, i.e. also with the erg. and dat., although this would require a more careful analysis.

morphology,[34] one such feature is the erg. pl. The isogloss erg. pl. *-ak* / *-ek* is close to the administrative frontier between Guipuscoa and the continental Basque Country, enters Navarre and crosses this territory through the middle. The traditional way of interpreting this was that both morphemes are different evolutions of a proto-form that was already formed in the common language (**mutil-ag(-e)-k* > west. *mutil-ak* / east. *mutil-ek*) (Michelena 1988 [1964b]: 10). But later research prefers to see a demonstrative determiner, rather than a bound morpheme, in the origin of the declension (Irigoyen 1981, Trask *HB*: 200, Manterola 2009b), although the details may differ as to which forms that approach can apply to. The erg. pl. form *mutil-ek* of the east could easily be explained as *mutil hek* → *mutil-ek*, but the erg. pl. *mutil-ak* of the west is not so easily derivable from a demonstrative (as these end in *-ek*). Perhaps west. erg. pl. *mutil-ak* is simply a syncretism with the abs. pl., whereas the east found an overt way of marking the erg. pl.[35]

An analogous geographic distribution of the vocalism (west. *-a-* / east. *-e-*) is found in finite verbal forms to which *-n* or *-la* is attached, as in west. *da-n* / east. *de-n* 'which is', both of which come from different developments of **da-e-n*, according to Michelena (1988 [1964b]: 10).

An important feature concerning non-finite verbs is the double way for making the prospective participle, with *-ko* in the west, *-(r)en* in the east, and a central region in which the tendency is for vowel-ending participles to take *-ko* and consonant-ending ones *-en* (*ikusi-ko*, but *erran-en*) (§8.3.3). These seem to be different elections for creating a periphrastic prospective. Both are formed upon the genitive, but while in the west the local genitive was chosen, in the east the possessive genitive was preferred. Another feature which behaves differently in the west and the east is the opposition of the verbal root and the pfv. participle in those verbs in which the opposition is possible (e.g. *has* vs. *has-i* 'to begin').

34 In the phonetic realm, one feature which is close to an A-A-A/B-B-B alignment is the phonetic or contextual palatalization. It is not exactly true that the further westwards we move, the more phonetic palatalizations we find, since e.g. in Soule the change *n*, *l* > *ñ*, *ll* / j _ V occurs, whereas in the Low Navarre and Labourd it does not, as we have seen in §2.2.2.6. But it is true that, in most of Navarre, Guipuscoa and Biscay (except its westernmost strip, within the Basque-speaking area), the palatalization is stronger than in the continent, as it includes the same change also after a syllabic *i*: *n*, *l* > *ñ*, *ll* / i, j _ V. In many peninsular regions, this palatalization process includes *t* (hence *t* > *tt*, *tx* / i, j _ V), or even processes such as *ild* > *illdd*, *ilt* > *illtt*, *ind* > *iñdd*, *int* > *iñtt* (cf. Zuazo 2008: 243 ff.). Besides, in B and Al *s*, *z* > *x* after an old diphthong in *j* (B. *axe* 'wind', vs. comm. *(h)aize*), and the same occurs with sibilant affricates (B. *atxur* 'hoe', vs. comm. *(h)aitzur*) (cf. *FHV*: 192). In these processes, the west must be innovating.

35 Similar problems are found when analyzing the dat. pl. west. *-ai* / east. *-ei* forms, although the existence of a far east. *-er* makes things more complicated here (§6.4.1.3).

The latter encroaches on the former in the west (west. *has-i naite(-ke)* vs. east. *has naite(-ke)* 'I can begin'). On this point, the west is clearly innovating (§8.3.1.3).

Two final features that I will mention among A-A-A/B-B-B alignments concern syntax. The first one is the *Transitive Objective Genitive* phenomenon (= TOG), which, in this case, opposes not the east to the west, but the continent (and some northern Navarrese varieties) to the peninsula. In the continent, direct objects of non-finite clauses whose verb is formed by the nominalizing suffix appear, or tend to appear, in the genitive (*ogi-aren eros-te-ra noa* 'I am going to buy the bread'). In peninsular Basque dialects, the direct object is expressed by the abs. in such cases, as in finite clauses, although in old —and not so old— texts there remain abundant traces of the phenomenon, even in the west. Here it is the continent that is conservative (cf. *HB*: 244-245). The second one is the productivity of the article. By and large, the further westwards we move, the more contexts the article has spread to, such as e.g. subject complements or indefinite direct objects (Sal. *Gu on-Ø gra* 'we are good', *Eper-Ø arpa-tzen zuen* 'he used to catch partridges'; cf. Arana 2006: 43, 187). However, in old west. texts the situation is much more similar to that of the east, and hence the spread of the article is an innovation (RS. *Gibel-eko on-Ø da-n-a* 'what is good for the stomach', Lcc. *eskalu-Ø ar-tu* 'to catch fish'; cf. §6.3.1).

2.2.3. How many Basque dialects are there?

A history of the discussions held by Basque dialectologists on this question would have no end. Besides, for the diachronist this issue is not decisive, since it consists in evaluating a current state of interrelations between speeches by equally considering all their elements, without trying to discriminate what is new and what is archaic. Hence, I shall mention only the essential points of the most influential proposals.

The first systematic attempt at a classification of the Basque dialects according to well-defined linguistic criteria was that of Prince Bonaparte.[36] But, actually, he made no fewer than four attempts. Bonaparte most probably made five trips to the Basque Country (in 1856, 1857, 1866, 1867 and 1869), in which he traveled all over the Basque-speaking territory (then more extensive than it is today), trying to learn as much as possible about each dialect. During those trips, he also formed an ample circle of collaborators, many of them priests, to whom he entrusted translations of biblical books, Christian doctrines and similar texts

36 I am following Pagola (1991: 115 ff.), in which there is also a synthesis (ibid.: 91 ff.) of the history of the question before Bonaparte, which we cannot enter here.

into each dialect. He thus resorted to two kinds of sources for dialectal research and taxonomization: a direct one, orally gathered by him in his trips; and the texts, in which he wanted the autochthonous variety to be reflected as closely as possible.

Bonaparte's first classification, conceived in ca. 1861-63 and expounded in his letters belonging to those years as well as in Bonaparte (1862), was rather tentative, and his criteria were variously defined depending on the dialect. In fact, he paid more attention to literary dialects, so he focused his interest on Biscayan (divided into two west. and east. subdialects), Guipuscoan and Labourdin. For these, he proposed a number of subdivisions. The others were the High Navarrese, the Low Navarrese and the Souletin, but, in the Prince's first classification, the subdivisions for these dialects are rather lax, according to Pagola (1991: 136). Interestingly, the parameters on which he based this classification were vowel harmony (*etxe-a* > *etxí-e*), yod-insertion (*berri-a* > *berri-ya*), the pronunciation of fricative sibilants and palatal affricates, and the erg. pl. realization (west. *gizon-ak* vs. east. *gizon-ek*), but he did not enter much further into morphological complexity.

Bonaparte's second classification is to be dated to 1864-65, and is described in a number of letters penned during those years. In this classification, he recognizes four literary dialects (B, G, L and S), and goes into detail as to how the Souletin and the High Navarrese territories are to be subdivided. The parameters employed this time comprehend more morphological aspects, such as the Nor auxiliaries (*naz* / *naiz* / *niz*), Nor-Nork auxiliaries (*dot* / *det* / *dut* / *düt*), the sociative case (*jaun-agaz* / *jaun-arekin* / *jaun-areki*), syntactic aspects such as relative clauses (*zein dan*, *den* / *zein baita*), lexical aspects (*esan* / *erran*), etc. In this classification, the territory of Soule is enlarged and moves to occupy the areas of Cize (southeast. Low Navarre) and Mixe (northeast. and north. Low Navarre), with Souletin now being subdivided into three subdialects: that of Cize, Mixe; that of Soule; and Salazarese, Roncalese (on the peninsular side). Low Navarrese, in turn, now moves to occupy an eastern part of Labourd, and also comprehends the peninsular Aescoan. The three LN. subdialects, then, are: that of Labourd (Hazparne, Hiriburu, Beskoitze), that of Low Navarre (Luzaide-Valcarlos) and that of Spain (Aescoa). He also tried to adjust some aspects in the west-center, and included the Navarrese Burunda and Arakil within the Guipuscoan dialect.

The third classification was made in 1866-68, after the trips of 1866 and 1867, and hence is much more refined in its detail and its argumentation as to

how subdialectal divisions are to be made.[37] Dialectal limits, however, do not considerably change with respect to the previous classification, except for the integration of Labourdin into Navarre, to constitute a dialect called *Navarrese-Labourdin*. Another terminological innovation is that Souletin now comes to be called *Navarrese-Souletin*. The parameters here are much more specific according to the limit to be established in each case, but verbal morphology gains in importance.

In fact, verbal morphology is the central axis of Bonaparte's fourth and definitive classification, as is to be expected from the title of the work in which it was presented: *Le verbe basque en tableaux* (1991 [1869]). Here the Labourdin dialect recovers its independence and its name, High Navarrese is split into Northern and Southern High Navarrese (both with dialectal status), and the Navarrese-Souletin of the third classification is split into Souletin (which comprehends Soule and the Roncal Valley in the peninsula) and Eastern Low Navarrese (which comprehends the Salazar Valley in the peninsula, the southeast. and north. Low Navarre and even sticks into a chunk of northeast. Labourd). The previous Low Navarrese is now called Western Low Navarrese (which comprehends Aescoan in the peninsula and also penetrates into Labourd). The Northern High Navarrese, in turn, snatches Arakil (but not Etxarri-Aranatz) and the northeastern end of Guipuscoa from the Guipuscoan dialect. Altogether then, they are 8 dialects, under which sub- and sub-sub-dialectal divisions are conscientiously worked out. These 8 dialects are also grouped into three main (supra-dialectal) zones: 1) B; 2) G, North.-HN, South.-HN, L; and 3) West.-LN, East.-LN, S.

Bonaparte made hardly any further changes to his fourth classification, but in 1881 he included Baztan within Labourdin (apud Yrizar 1991-2008: Labourdin, 2), a possibility already hinted at in *Le verbe basque en tableaux*. This classification was accepted by Arturo Campión (1977 [1884]), and in fact, has been accepted until recently (until now in philological practice), although each author has made the retouches that (s)he has considered convenient. Thus, in the Introduction of his *Diccionario vasco-español-francés*, Resurrección Mª de Azkue (1969 [1905-06]: XXVI-XXX) pointed out that he would follow Bonaparte's classification, but with the following changes: 1) Roncalese now becomes an independent dialect; 2) North.-HN and South.-HN go on to be subdialects of a single dialect, HN; and 3) West.-LN and East.-LN are also reduced to a single dialect, LN, of which both are subdialects. In his view (ibid.: XXVII), if one analyzes "the declension, the derivation, the word compounding, the conjugation,

37 It is expounded in Bonaparte (1866) and also in an undated work with a title which is several lines long but begins with *Phonologie de la langue basque dans tout ses dialectes*... (apud Pagola 1991: 145).

etc." [translation mine, MMA], one can see that Roncalese and Souletin differ from each other at least as much as Guipuscoan and Labourdin do, and that the differences between HN and LN do not suffice to split each one into two dialects. 7 dialects, then, are the ones given in Azkue's classification.

Luis Michelena (*FHV*: 41-42) raised this number to 12 dialects, assigning that status to Al, Aesc, Sal and R, in addition to the ones proposed by Bonaparte. However, rather than to propose a new dialectal classification, Michelena's intention was simply to provide a more useful terminology for handling phonetic-phonological issues, which obviously constituted his main concern in the *Fonética Histórica Vasca*. He admitted that Aesc, Sal and R were morphologically close to West.-LN, East.-LN and S (in fact, the Prince's most important criteria had been morphological), but noted that phonetically they differed.

Pedro Yrizar (1981, 1991-2008), the Basque dialectologist who, after Azkue and Bonaparte, has researched Basque dialects in the greatest depth, operated with 9 dialects: to the 8 proposed by Bonaparte, he added R (like Azkue). He also included Baztan within L, a possibility already contemplated by Bonaparte himself in *Le verbe basque en tableaux*, and preferred by him in 1881.

Recently, Zuazo (1998b, 2008) has come up with a different classification. He has noticed that Bonaparte's classification did not distinguish between innovations, archaisms and elections made by each dialect. Instead, Zuazo's proposal (1998b: 192 ff.) is to delimit innovations and to draw a classification based on them, leaving aside archaisms such as e.g. the maintenance of voiceless stops after *n*, *l* in R and S (as in S. *igánte* 'Sunday', *sükálte* 'kitchen', vs. SB. *igande*, *sukalde*). Besides, for innovations to be taken into consideration as a delimiting element, they have to fulfill three criteria to the greatest extent possible: a) to be widespread throughout the Basque-speaking geography; b) to be productive in terms of frequency and regularity; and c) to be alive in today's Basque. Basing on these criteria, he has renamed (some of) the classical terms for Basque dialects, and reshaped their classification, proposing the following dialects (ibid.: 228-229): 1) Western (the former B, including what remains of Al); 2) Central (the former G, but also comprehending some western North.-HN. areas like the two Basaburuas, Araitz and Larraun, and the Guipuscoan territory which belonged to North.-HN); 3) Navarrese (the former North.-HN, but also including the whole territory up to Aescoa, and Baztan); 4) Eastern Navarrese (Sal and R, now defunct); 5) Souletin (including the whole territory of Soule, although the northwest. part is very close to the variety of Mixe, in the Low Navarre); and 6) Navarrese-Labourdin (which includes all the other continental varieties, and also comprehends Urdazubi and Zugarramurdi, and probably Luzaide-Valcarlos). 6 dialects, then, are to be found in Zuazo's classification.

2.3. The origin of Basque dialects

Now that we have very briefly described the Basque dialects and some of their most representative features, we can turn our attention to the question of how they have developed diachronically, and even how they may have arisen. The study of this issue implies, on one hand, the need to interpret the historical dialectal diversity and, on the other hand, the need to resort to prehistoric sources (roughly, previous to 1500), thus seeking an interpretation which will make the data of those sources fit in the best way possible with historical facts. This issue will be brought up by briefly summarizing some of the contributions which have tackled it in one way or another.

2.3.1. The secular view: tribes of Antiquity, Church dioceses, historical dialects

After Bonaparte, Basque dialectology had a reasonable and useful classification of Basque dialects. But this was a purely synchronic issue. Even accepting that classification, the question remained as to how they were formed.[38]

In the 20th century, Arturo Campión (apud Pagola 1991: 81) was the first to notice that, at least as far as the peninsular Basque Country was concerned, there was a certain correspondence between the divisions of Basque dialects as established by Bonaparte and the divisions of the tribes which occupied the historically Basque territory as traced by the Greek-Egyptian geographer Ptolemy in the 2nd century AD. Before we go further into this analysis, we can examine Map 2.4, in which the tribes of the central north of the Iberian Peninsula appear according to Ptolemy's description and assigned limits.

Beginning from the west, the Autrigones would be the tribe which occupied the territory from the river Asón to the river Nervión, comprehending the west. part of Biscay, a west. strip of Alava and the northeast. part of Burgos. The tribe of the Caristii, in turn, would occupy the territory from the Nervión to the river Deba, thus comprehending what is roughly the historically Basque-speaking Biscay, but also the west. basin of the Deba Valley, which in the historical period administratively belongs to Guipuscoa. It would also stretch into Alava, occupying the northwest. part. The Varduli would occupy the territory from the Deba up to the limit with the Vascones, and they would also stretch into Alava, occupying almost the entire east. half. To the south of the Autrigones, the Caristii, the Varduli and partially the Vascones, and occupying a region which now belongs to La Rioja, we would have the tribe of the Berones, to which, in the 1st

38 I am following Pagola (1991: 77 ff.) and Zuazo (2010a: 139 ff.) in §2.3.1, and Pagola (1991: 172-176) and Michelena (1988 [1964b]: 19-22) in §2.3.2.

c. BC, Strabo assigned a Celtic filiation. To the east of the Varduli and the Berones, in what is roughly the historical Navarre, but sticking into the north-east. area of Guipuscoa and, on the east. side, occupying an ample strip of the historical Aragón, we would have the Vascones.

Map 2.4. Tribes of Antiquity according to Ptolemy (2nd c. AD)

Although the reliability of Ptolemy's classification has been much discussed due to its numerous differences with respect to the classification made by some other earlier Greek-Roman authors, current research tends to accept it (Gorrochategui 2002: 105-106). Now, to continue with Campión's reasoning, Caro-Baroja (1990 [1945]: 34) observed that the limit traced by Ptolemy between the Vascones and Varduli corresponded to the linguistic limit between North.-HN and G, not to the current administrative one between Navarre and Guipuscoa. Similarly, the limit between the Varduli and the Caristii did not correspond to the administrative limit between Guipuscoa and Biscay, but was much closer to the linguistic limit between G and B, with the Caristii occupying a strip of the historical Guipuscoa. He concluded: "It seems licit, then, to suppose that in the Roman period the Vascones, the Varduli and the Caristii already had their own dialects within a common language" [translation mine, MMA]. Zuazo (2010a: 149) observes that this secular view is nowadays rejected by most linguists (within the domain of Bascology), but is still accepted by most dialectologists (but not by himself, cf. §2.3.4).

To this correspondence 'tribes of Antiquity → historical Basque dialects', Serapio Múgica (1914-17: 192-193) —apparently inspired by Carmelo de Echegaray— had already added another element which could be interpreted as a chronological bridge between its two poles: the medieval church dioceses. Thus, the North.-HN. dialect would have corresponded to the Diocese of Bayonne, the G. dialect to the Diocese of Pamplona, and the B. dialect to the Diocese of Calahorra (into which the earlier Diocese of Alava was integrated in the 11th century). With different nuances, some important historians like Menéndez-Pidal and Andrés de Mañaricúa have defended the triple correspondence 'tribes of Antiquity → medieval church dioceses → historical Basque dialects' (apud Pagola 1991: 83-84).

However, Zuazo (2010a: 140-145) has objected that the theory of such a continuity does not hold up. To begin with, the largest part of Guipuscoa and Navarre were within the Diocese of Pamplona, which is incoherent with the fact that Guipuscoa corresponded to the Varduli in Ptolemy's classification, and Navarre to the Vascones. Besides, citing Barrena (1989: 283 ff., 342-343), he notes that an important part of Guipuscoa formerly belonged to the Diocese of Bayonne, and that it was only in a later period that it was integrated into the Diocese of Pamplona, with the exception of Oiartzun, Irun, Hondarribia and some other towns of northeast. Guipuscoa. This is unexpected if we bear in mind that this is, at least in theory, a North.-HN. region (and hence the one linguistically closest to the surroundings of Pamplona-Iruñea in Antiquity). Church dioceses and linguistic divisions do not fit with each other in the continental Basque Country either. Here, there were three church dioceses: Bayonne, Dax and Oloron. There are several ways of classifying the continental dialects (see §2.2.3), but not one shows any correspondence between dioceses and dialects, at least as far as Labourd and Low Navarre are concerned. For instance, while the Low Navarrese area of Oztibarre was included within the Diocese of Dax and the (also Low Navarrese) area of Cize within the Diocese of Bayonne, the varieties of Oztibarre and Cize are very close to one another.

The correspondence between the tribes of Antiquity and historical dialects is also problematic. As we have seen in §2.2.2, a number of important structural features typical of west. Basque, such as the process *-a-a* > *-ea*, stretch eastwards, along the southern axis from Vitoria-Gasteiz to Pamplona-Iruñea, into the Navarrese area of Burunda, which is unexpected according to the vertical positioning of Ptolemy's tribes.

2.3.2. Lacombe (1952 [1924]), Uhlenbeck (1947 [1942]). Biscayan vs. the other dialects

We have seen in §2.2.2.1 that, as a dialect, Biscayan is very differentiated from the rest. No Bascologist has escaped this impression, although the interpretations have been diverse. According to Lacombe (1952 [1924]), the differences between Biscayan and the rest of the dialects are so great that we can group all Basque dialects into two groups: 1) Biscayan; and 2) the central-eastern group (all the other dialects). The diachronic implication of this would be that, when the common language split, it branched into two groups, one being the ancestor of historical Biscayan, and the other being the ancestor of all the other historical dialects. Lacombe thus proposed what we could call a kind of *Biscayan-Basque*, somewhat reminiscent of Sturtevant's Indo-Hittite.

This idea was not entirely new, as it had already appeared in Campión, Broussain, and Duvoisin. Julio de Urquijo also adhered to the same view in 1929 (apud Pagola 1991: 173). But Uhlenbeck (1947 [1942]) went further and proposed that Biscayan and the ancestor of all the other Basque dialects had, in origin, been different languages to be located in the western Pyrenees, and that at some point they had come into contact with each other, mingling but still preserving some traces of their different origins.

None of these views has resisted criticism. Let it suffice to cite Tovar (1959: 146 ff.), Michelena (1988 [1964b]: 19-22) and Lakarra (1986), and refer to §2.2.2.1. There, it was shown that at least many of the distinctive features of Biscayan are a consequence of the diverging vicissitudes from a pool common to all dialects. Even structural features such as the presence of the root **gin* in pot., sbjv. and imp. forms, which apparently point to elections made by the dialect at a very early point in the fragmentation process, have been shown by Lakarra (1996a: 184) to have most probably occurred relatively late. In fact, Michelena's (1988 [1964b]: 16; citing Rodríguez-Adrados) observation according to which the differentiating features that model dialects can be classified into three groups, namely archaisms, elections and innovations, is also applicable to B: e.g. archaic inessive in *-n* (*itaurre-n*), election of **duts* as the root of Nor-Nori-Nork auxiliaries, abl. in *-rean*.

2.3.3. Michelena's (1987 [1981a]) éuscaro *or Common Basque*

Michelena (1987 [1981a]) is an attempt to explain how the diversity presented by Basque dialects in the historical period may have arisen. Throughout this work, he moves from the linguistic issues to historical ones. With regards to the former, he insists, partially repeating arguments of Michelena (1988 [1964b]),

that B is a branch of Basque, at the same level as each one of the rest of historical dialects. But, taking B as a starting point, he argues that the diversity of features present throughout all the dialects is neither so old nor so deep as to bring the origin of the dialects back to an extremely old period. Biscayan in particular is known to have innovated in several features in the historical period. Grammatical examples of this type of development can be found in §2.2.2.1. Regarding the lexicon, he admits that a number of B. words may sound rather peculiar in central regions, but not more so than their east. equivalents (compare west. *jazo* 'to happen' vs. east. *agitu*, *gertatu* being more widespread than both; or west. *arerio* 'enemy' vs. east. *izterbegi*, *etsai* being general). In the realm of phonology, B or Western Basque in general has a number of peculiar features, such as *e* > *a* / _ -rr- in a number of lexical items (west. *barri* 'new' vs. *berri*), west. *baltz* 'black' vs. *beltz*, etc., but these are independent innovations.

Of extreme importance is the fact that all dialects, including B, participate in the set of procedures by means of which the earliest loan words are borrowed, such as the intervocalic dropping of *-n-* (Lat. *annōna* > *an(h)oa* 'daily ration'), *-l-* > *-r-* between vowels (Lat. *colu* > *goru* 'spinning wheel'), the voicing of initial stops (Lat. *castellu* > *gaztelu* 'castle'), etc. Also, all dialects partake in the processes which are undergone by the end of the first member in composition and derivation junctures, such as *ardo* 'wine' → *ardan-tza* 'vineyard', *gazta* 'cheese' → *gaztan-bide* 'street of the cheeses', *beso* 'arm' → *besa-pe* 'armpit', etc.

There must have existed, then, an *éuscaro* or a common language from which the historical diversity derives. But, where is it to be chronologically located? Michelena's approximate account of the facts is as follows (and we are moving now to the aforementioned historical issues) (cf. ibid.: 52-55). Romanization brought about, in the realm of language, a drastic reduction of the Basque-speaking area, just as, in fact, the pre-Latin Indo-Europeization had already done. The remaining focus must have attracted, especially in periods of violence and insecurity, ill-adjusted populations from areas which were far from that central nucleus, and who consequently spoke diverse dialects and varieties (of a Basque previous to that common language). This process may have lasted until the 2nd c. AD, but probably quite a lot longer.

That nuclear region, which probably covered, according to Michelena, the surroundings of the western Pyrenees on both sides and the Atlantic area of the historical Basque Autonomous Community, kept its old gentile social organization, as opposed to the new urban order based on large-scale landholding imposed by Rome in the surrounding area. Of course, it is also possible that stocks of Basque or Bascoid peoples could have survived further away from that con-

servative nucleus, as in the case of Pallars and the High Ribagorza (both in Lérida), which may have been Bascoid until the 10th-11th centuries.

That gentile or tribal society was reinforced from the moment that the imperial administration began to malfunction (quite soon, in fact). And, crucially, according to Michelena (citing Maluquer), although Roman influence may have been scarce in socio-economic as well as in spiritual aspects, the administrative structure which surrounded the Bascoid nuclear region may have decisively contributed to sparking its sense of unity. This sense of unity probably increased during the period of Visigothic domination, now openly hostile. Along with other factors, this sense of unity is one of the elements which ultimately gave rise to the Kingdom of Pamplona (from roughly 800 onwards). Especially the 5th century and part of the 6th one were, in all probability, a period of absence of external domination, and hence also of social reorganization and change. The importance of kinship networks, so typical in gentile societies, probably laxed, and the social hierarchization increased. This provoked interchange at all levels, and a tendency to a greater unity and centralization. The effect of this upon the language was that common forms were favored to the detriment of regional or local forms. A Bascoid *koiné*, then, was formed.

The two key factors of the spread of this *koiné*, Michelena (ibid.: 54-55) concludes, were the introduction of Christianity and the establishment —and consequent influence upon neighboring regions— of the Kingdom of Pamplona (which in turn was probably a factor of the consolidation of Christianity). Michelena further notes that Biscay was the territory most reluctant to the introduction of Christianity and to the influence of Pamplona, and was archaic in many socio-cultural aspects. Since that linguistic *koiné* irradiated from Pamplona-Iruñea, this might also explain the (relative) singularity of Biscayan as a dialect.

2.3.4. Zuazo's (2010a) innovating foci

Zuazo (2010a) is an essay which, insofar as the historical aspects of the question are concerned, can be considered as a specification of how the process of expansion of Michelena's *éuscaro* or Basque *koiné* may have occurred. In fact, (2010a: 139-164) is an approach to the problem of the antiquity of Basque dialects, and their relation to Ptolemy's tribes and the medieval church dioceses, more or less as has been outlined in §2.3.1. He further lists the linguistic arguments which support Michelena's view that Basque dialects must have arisen in a period later than Antiquity, and approves that conclusion.

On the following pages, Zuazo (2010a: 165-193) proposes that, in the spreading process of that *koiné* or common language, the innovating foci were

the following: 1) Pamplona-Iruñea; 2) Vitoria-Gasteiz; 3) Center of Biscay; 4) Soule and the Low Navarre; and 5) the area of Beterri (in Guipuscoa). All of these foci are urban centers (even Soule and the Low Navarre had some urban points). He also suggests that Pamplona-Iruñea was the first innovating focus, Vitoria-Gasteiz the second, and the rest came somewhat later.

Pamplona-Iruñea was the most important Basque town in Antiquity. As early as in the 6th century it was a Church Diocese, and it was crossed by the Roman road which went from Astorga to Bordeaux. That urban vigor was reinforced in the Middle Ages by St. James' Way. It is then conceivable that some innovations which affected all dialects had Pamplona-Iruñea as the irradiating focus. Some of these are the initial voicing of stops, *-n-* > *-Ø-* and *-l-* > *-r-*, etc., between vowels, the processes undergone by the first member in composition and derivation, the allocutive treatment (*naiz* → *nau-k/n*; cf. §9.6.2.1.2), and the creation of a 2nd p. sg. of respect from the standard plural (cf. §7.3.1 and §9.3.2), completed with the exception of some agreement markers in the west. Some other innovations which may have had their irradiating point in Pamplona-Iruñea, but have not reached all areas and are perhaps to be dated later, are the following: the morpho-phonological process *-a-a* > *-a* (*alaba-* + *-a* > *alaba*; i.e., no morpho-phonological process takes place, as opposed to *-a-a* > west. *-ea* / Sal. *-ara* / S. R. *-á*); the morpheme *-te* as an agreement marker of 2nd p. pl. abs. (*zare-te* 'you (pl.) are', as opposed to west. *zarie* < **zara-e* / east. *zidie* < **zara-de-e*); and the emergence of *-tzen* as the main ipfv. suffix of non-finite verbs (cf. §8.3.2.2.2).[39]

In turn, Vitoria-Gasteiz was the most important urban center of the whole western area from the middle of the 11th century onwards. At the end of that century, the city was partially surrounded by a wall, and the Diocese of Alava, later integrated into that of Calahorra, played an important role. This suggests that Vitoria-Gasteiz was an irradiating focus, and in fact only this can explain, according to Zuazo, the fact that some important features like the *-a-a* > *-ea* process of noun morphology or the vocalism of Nor-Nork auxiliaries such as *dot* 'I have' or *dau* '(s)he has' spread not only throughout the entire B. and Al. dialects, but also into regions of the G. dialect, and even into Navarre in areas like Burunda, Amescoa and Lana. Zuazo further proposes 13 features which might have had their irradiating point in Vitoria-Gasteiz. Some of these features are the palatalization of *z*, *tz* into *x*, *tx* after *i* with the subsequent disappearance of that *i* (*haitz* > *atx* 'crag'), the directional *-rutz* (cf. §6.5.1.1), the abl. *-rean*, the generalization of the root *eutsi* for Nor-Nori-Nork auxiliaries, and —inspired by

39 In fact, many of the A-B-B-B-A alignments mentioned in §2.2.2.10 could also belong here. Obviously, the feature which spread from Pamplona-Iruñea in such cases was B.

Urgell (2006)— the extension of the suffix *-etan* as ipfv. and verbal nominalizer in verbs ending with *-adu*, *-atu*, *-idu*, *-e* (cf. §8.3.2.2.3).

A later focus of innovations, which somehow continued the innovating processes initiated in Vitoria-Gasteiz, was the central area of Biscay, perhaps formed by urban points such as Durango, Amorebieta-Zornotza, Gernika and Bermeo. These constituted an axis which linked Vitoria-Gasteiz with the coast. According to Zuazo, the Basque spoken in Bilbao is archaic in several aspects, but Bilbao was never an innovating focus. Some of the features which may have arisen in the central area of Biscay are the diminutive suffix *-txu* (comm. *-txo*), *-iño-* instead of *-io-* in the root *-io-* 'to say' (*dio* '(s)he says', B. *diño*), and the anteposition of the demonstrative to the noun (§7.2.4). Perhaps more modern are the tendency of the diphthong *-au-* to become *-eu-* (*gaur* 'today', B. *geur*), the palatalization of *z* after *i* (*bizi* 'to live', B. *bixi*), and the extreme productivity of the abs. pluralizer *-z* (as in *n-it-uen* 'I had them', B. *neba-z-an*; cf. §2.2.2.1).

In the northeast. part of the Basque-speaking territory, another focus which must have irradiated innovations, here again perhaps continuing those of Pamplona-Iruñea, is to be located in Soule and the Low Navarre, most probably in their urban centers, such as Maule, Atharratze (Soule), Donibane Garazi and Donapaleu (Low Navarre). Zuazo distinguishes two types of innovations here: those corresponding to the entire northeast, and those which affected only Soule and perhaps the neighboring Roncal. Among the former, he includes the morpho-phonological process *-u-a/e* > *-uya/e* > *-ia/e* (*esku-a* > Dch. *eskuya*, S. *eskia*), the reduction of the vocalism in forms such as *naiz* > *niz* 'I am', *zaitut* > *zitut* 'I have you (sg.)', *nauzu* > *nuzu* 'you (sg.) have me', the dat. pl. *-er*, and the allocutive treatment with *zuka* (cf. §9.6.3). Among the latter, we would have the morpho-phonological process *-a-a* > *-á*, the change *-au-* > *-ai-* in some contexts, and the all. in *-ala(t)* (Bp. *bekhati-ala* 'to the sin').

Finally, Zuazo considers that the area of Beterri, which stretches from Donostia-San Sebastián to Tolosa and contains some other important urban centers such as Hernani and Andoain, may have been another innovating focus, related to the features which define Guipuscoan, such as the *-e-* of Nor-Nork auxiliaries, the *-e-* of Nor auxiliaries like *zera* and *gera* (cf. §2.2.2.2), and the special finite forms of the verb *joan* 'to go' like *nijoa* 'I go', *dijoa* '(s)he goes', (vs. *noa*, *doa*) etc. The prominence of this focus is to be dated perhaps later than the others.

Zuazo concludes his essay by considering the western dialect to be the most diverging one, and also the one which first broke away from the rest.

2.3.5. Camino (2011), Lakarra (2011f)

In two papers to be published, Camino (2011) and Lakarra (2011f) have defended a view which, as regards the topic of this section (how the dialectal diversity came about from a common trunk, and which splits took place first), is contrary to the one defended in Zuazo (2010a).

Camino (2011) makes a thorough analysis of both linguistic and extralinguistic aspects of the eastern area of the Basque language. He calls into question Zuazo's suggestion that innovations have irradiated from urban centers (Camino sticks, however, to the situation in the east). According to him, even if Maule and Atharratze have been market towns, in a territory like Soule, where the pastoral economy has traditionally been so strong, it is extremely doubtful that such centers have played a significant linguistic role in linguistic innovations beyond *leveling* processes, and in fact there is no evidence that linguistic features which have originated there have spread to the whole eastern territory. Camino (2011) also recalls that, by 1820, the four most populated towns were Barkoxe, Sohüta, Santa Grazi and Montori (which were shepherd communities), Maule and Atharratze not being among them.

As for purely linguistic aspects, Camino (2011) analyzes 140 linguistic features, and notes that in 91 of them the Soule and the Roncal have gone hand-in-hand throughout the prehistory and the history of the language. They also share 27 archaisms, and, in 10 further features, they have proven impenetrable to innovations coming from the west (whether on the continental or the peninsular side). They also share 32 innovations, 17 of them exclusive to the area shared by both. While the latter argument supports the historical unity of S and R, the sharing of so many archaisms (some of them structural, cf. below) points to an early split from the common language of both.

Lakarra (2011f), in turn, analyzes, and to some degree revises, the implications of Michelena's (1987 [1981a]) *éuscaro* or Common Basque in the light of the advances that Basque diachronic linguistics has made in recent years. His analysis covers every realm of the language, including substantial argumentations about the relative chronology of phonetic changes such as **b- > Ø-*, *k- > Ø-*, /p-, (t-,) k-/ > /b-, (d-,) g-/, the metathesis *a-o > o-a*, **eV- > jV-*, the Grassmannian dissimilation **h-h > Ø-h*, **-r > -h*, **h_3 > h_1*, *-VØV- > -VgV-*, etc., some of which we did not know some few years ago. He also analyzes word structure phenomena, the prefixing morphology of non-finite verbs, morphological issues such as the roots of auxiliaries, etc.

In what concerns us here, Lakarra opposes Zuazo's view that it was the west. dialect —or in a deeper sense, Western Basque— that branched off earliest. He notes that, however special or idiosyncratic Western Basque may appear

in the historical period, even its earliest features are not as old as some others that can be found in the east. We have seen some of these aspects, such as the process *-a-a* > *-ea*, the generalization of **gin* as the pot.-sbjv.-imp. root of Nor-Nori-Nork auxiliaries, etc., in §2.2.2.1. Instead, he prefers to look to the north-east. corner of the Basque Country to find splits which can be brought back to the earliest phases of the fragmentation of the common language.

As I said above (§2.2.2.9), unlike the rest of the dialects, S and R do not undergo the change *k*, *t* > *g*, *d* / n, l _ . Although the usual interpretation is that a voicing rule is applied in this context, but that that rule did not reach S and R, Michelena (*FHV*: 355) hinted at the possibility that what really happened was that in the old language the opposition *fortis / lenis* was neutralized in that position, so that it was realized as *lenis* (or voiced), just as in initial position (§4.4.2.2). Thus, words like *alt(h)e* would be complex words which would have to be analyzed as *al-t(h)e*, S and R being the dialects which first abandoned this practice. Lakarra notes that this approach to the phenomenon is more coherent with the *fortis / lenis* opposition which is reconstructible for PB, as well as with Trask's Law (1995b) according to which, when we have a dialectal alternation in a suffix of the type -TV / -DV (where T = voiceless stop, D = voiced stop), the archaic variant is -DV (as is the case with the verbal pluralizer *-de* / *-te*, the adjectival *-di / -ti*, etc.; cf. §9.3.2 and §6.4.2.3, respectively). A further argument which would support this approach is that, in all of the dialects, after /n, l/ (but not after /r/!), sibilants have been exclusively fricative until very recently (and still are in some areas, as in the Burunda; cf. Zuazo 2010b: 68), when a fricative → affricate / n, l _ rule has applied. While this appears to be an innovation of S and R, an innovation of such a nature is to be dated to such an early chronological point that it supports a branching of S and R on one side and the remaining dialects on the other.

Another argument in favor of a split S and R (or Eastern Basque) vs. Central and Western Basque in the earliest phase would be the change *au* > *ai* in S and R (§2.2.2.9). Only after this had been accomplished could the change /*wh/ > /f/ occur, from which **gau-hari* > *afari* 'dinner' in all dialects except S and R (and Mixain and Sal, according to the *OEH*), where we find S. *aihari* and R. *aigari*. Since *afari* is (almost) general outside the area where *au* > *ai* applies, this must entail an earlier split of the varieties which undergo this change. He compares *afari* (~ S. *aihari*, R. *aigari*) to *jan(h)ari* 'food', *ba(ra)zkari* 'lunch' (< *baratze* + *hari*) and *askari* 'breakfast; afternoon snack' (< **arrats* + *hari*), which are general or in any case are present both within and outside the area of S and R (see *OEH*).

Therefore, unlike Zuazo, Lakarra (2011f) defends an earlier split from CB of the eastern branch, which is represented in Figure 2.1, in which dialectal splits are binarily conceived.[40]

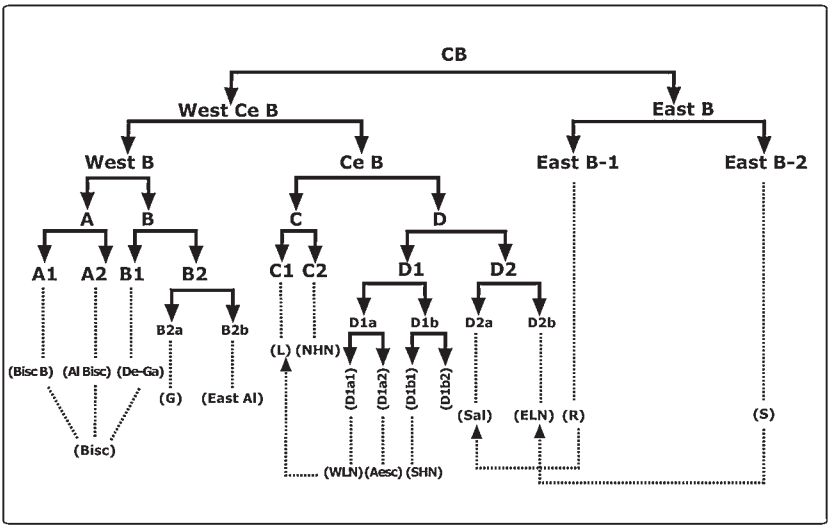


Figure 2.1. Diachronic dialectal diversification hinted at in Lakarra (2011f)

In Figure 2.1, the items in parentheses are terms used in Bonaparte's dialectology or in the dialectology developed after him. Dotted lines represent dialectal configurations developed after the medieval period, and dotted lines or arrows after items in parentheses represent historical convergence processes developed due to the influence of one dialect upon another. It must be stressed, however, that the argumentation in Lakarra (2011f) sticks to the branching of the first node. The branching of the following nodes sketched in Figure 2.1 is simply a plausible development. In the future, diachronic dialectology can brush up the details, or even propose alternative developments.

40 *Legenda*: CB = Common Basque; West Ce B = Western and Central Basque; East B = Eastern Basque; West B = Western Basque; Ce B = Central Basque; Bisc B = Biscayan Basque; Al Bisc = Alavese Biscayan; De-Ga = Deba-Gasteiz; Bisc = Biscayan; G = Guipuscoan; East Al = Eastern Alavese; L = Labourdin; NHN = Northern High Navarrese; WLN = Western Low Navarrese; Aesc = Aescoan; SHN = Southern High Navarrese; Sal = Salazarese; ELN = Eastern Low Navarrese; R = Roncalese; S = Souletin.

2.4. Common Basque on the map

A final question concerns the location of this Common Basque, spoken in ca. the 5th-6th centuries, on the map. This brings us to the question of where Basque or Bascoid languages were spoken in Antiquity. Since the aim of this volume is to sketch a historical grammar of Basque, regardless of where this was spoken, this question is not crucial. Hence, it will be addressed in very general terms, just mentioning the most important current trends of research which can be found in the most recent literature.[41]

In the literature of recent years, views have ranged from those who believe that Basque(-Aquitanian) was spoken from the western border of the Caristii (the Nervión River) to the high basin of the Garonne, like Michelena (1987 [1981a], 1987 [1982a], 1987 [1982b]) and Gorrochategui (2004, 2007-08, 2009a, 2009b), to those who believe that the bulk of the Basque-Aquitanian population is to be located within Aquitania, like Villar, and Abaitua & Unzueta (2011). In particular, Villar (in Villar & Prósper 2005; part 3 of the monograph) has scrutinized the toponyms attested in Greek-Roman authors of Antiquity south of the Pyrenees, arguing that virtually all of them are non-Basque-Aquitanian. However, as Gorrochategui (2009b) has objected, toponyms are not a consistently reliable means to know about the language of the corresponding population. As he notes, the application of Villar's arguments and methodology to the corresponding place names of the continental side —to which Villar ascribes the original location of the Basque language— would lead us to the conclusion that that region was not Basque-speaking either, since Aquitanian river names like *Aturi*, *Garonna*, *Leyre*, or Aquitanian tribe names like *Vasates*, *Ausci*, *Tarbelli*, *Sotiates*, *Tarusates*, etc., do not seem to be any more Basque than the peninsular ones studied by Villar.

The study of indigenous anthroponyms and theonyms —attested in inscriptions of the period, which are more direct sources— made by Villar is not so detailed (cf. Villar & Prósper 2005: 497-502). He rightly notes that, in the historical Basque-speaking territories to the south of the Pyrenees, most indigenous anthroponyms and theonyms recorded in Antiquity for the Basque territory are non-Basque-Aquitanian. In fact, the greatest number of Aquitanian names, and particularly the great majority of those with a correspondence in historical Basque, are from Aquitania. But here we ought to consider in a different fashion the territory of the Vascones —and their surroundings to the south and east— and that of the Varduli and the Caristii (cf. Map 2.4). Within the former, Villar

41 For a more detailed summary in English, cf. Trask (*HB*: 9-12, 35 ff.), and especially Gorrochategui (1995a), although obviously this does not include the data discovered later.

counts as Basque-Aquitanian only those with a clear correspondence in the historical Basque lexicon (e.g. *Umme.sahar*, which actually corresponds to two words in the Basque lexicon, *ume* 'child' and *zahar* 'old') and those which do not have a clear correspondence in historical Basque but which have a correspondence in the Aquitanian inscriptions of the High Garonne and its surroundings (e.g. *Narhun-ges-*, similar to the *Narhon-s-* of the Hautes-Pyrénées). In all, Villar counts 10 Basque-Aquitanian person names in this area (which stretches up to Botorrita, in Zaragoza), opposed to 29 Iberian names (the majority of them from the Bronze of Ascoli) and 9 Indo-European names (ibid.: 499).

However, according to Gorrochategui (1995a: 52 ff., 2004: 117-121, 2009a, 2009b), such a count is misleading, as there are inscriptions with names which appear to be Iberian-Basque combinations (e.g. *Agir-seni-*), or which are Iberian but show phonetic and/or morpho-phonological features which are exclusive of Aquitanian and Old Basque, such as the aspiration and the gemination of the stem-final segment in *Urchatetell-i* or the affrication of the stem-final sibilant in *Ordunets-i* (in both cases, *-i* corresponds to the Latin ending). These are all included by Villar within the group of non-Basque-Aquitanian, which does not reflect the complexity of all linguistic phenomena. A one-by-one analysis is also problematic when considering some anthroponyms of the Bronze of Ascoli, as in them there are Bascoid elements. This would be the case, for instance, of *Arbiscar*, whose second element offers a plausible comparison with Basque *bizkar* 'back, hill'. While the comparison of these elements found in the territory of the Vascones —or its surroundings to the south and east— is most often not as transparent as many onomastic items found in the High Garonne and its surroundings,[42] an analysis of them as simply non-Basque-Aquitanian is an oversimplification of the facts.

Gorrochategui (2004: 118, 2009b: 76) also recalls that, in the area surrounding Estella-Lizarra (west. Navarre), where most indigenous anthroponymic items are Celtic, we find theonyms like *Loxa* and *Selats*, with plausible Basque correspondences. According to him, this appears to point to the well-known phenomenon whereby person names —more subject to fashions— are usually given up more readily than beliefs and god/divinity names. Hence, here we

42 The exceptions are the aforementioned *Umme* and *sahar* (inscription of Lerga, Navarre), which most probably refer to a single person but contain two different words in Basque, and perhaps the theonym *Losa* ~ *Loxa* (recorded four times throughout Navarre), which offers a comparison with Basq. *lotsa* 'shame, fear'. To these, we can now add the person name *Sesen-co* (a *cognomen*), which offers a crystal-clear comparison with Basq. *zezenko* 'young bull' (*zezen* 'bull' + dim. suffix *-ko*, abundantly present in the inscriptions of the High Garonne), and which was found some years ago in the highlands of Soria (quite a long way away from Navarre, to the southwest).

might have a Basque or Basque-Aquitanian autochthonous population which would have celticized its anthroponymic system.

To sum up, the Basque-Aquitanian element among the Vascones and their southeast. surroundings is far from being anecdotal. The question as to why it is not as copious and transparent as in the High Garonne region, however, is still open. It should be born in mind that Ptolemy's classification of the tribes was not based on linguistic criteria. In the case of the Vascones, under this tribal name there most probably underlay populations of different filiations and different socio-economic models, the area closest to the Pyrenees having an economy largely based on pasturage with seasonal migration of livestock, and that closest to the Ebro Basin based on sedentary agriculture (of large estate production from the arrival of the Romans on), probably with both models having commercial interchange. The orography of Navarre itself does not leave us much room to expect a scenario very different from this.[43] This does not necessarily imply a straightforward and exact correspondence between language and socio-economic model. It is conceivable, for instance, that Basque-Aquitanian populations had become sedentary in southern areas, closer to the Ebro Basin.

The territories of the Caristii and Varduli pose special problems. In Alava (cf. Gorrochategui 1995a: 51-52), there are abundant indigenous person names, almost all of which are Celtic.[44] An eyebrow-raising fact is that some of these

43 It might be interesting to introduce into the discussion the question as to how such socio-economic models may have had some influence upon language replacement, language change, language extension, language similarity throughout the whole territory with Bascoid onomastic elements, etc. As Mallory (1989: 146) points out: "It is also highly probable that the economic system of the speakers of a given language will play a significant role in the size of area in which relative uniformity will be maintained. Mobile subsistence economies such as hunter-gatherers, or more certainly pastoral nomads, frequently retain linguistic uniformity over a wider area than is typically found among agriculturalists among whom long-term village settlement will probably promote regional developments". Camino (1997: 97, citing Polák 1963) recalls the case of Albanian dialects, the relative unity of which has been favored by the importance of nomadism in the economic system of their speakers. What is recorded —virtually until today— for the Basque Country, and is expected to have existed along the western half of the Pyrenees —up to the Garonne— at the time of the arrival of the Romans, is not exactly pastoral nomadism, but seasonal migration of livestock. But the parameter is worth considering, as I see it.

44 A possible exception with a Basque-Aquitanian correspondence is *Seni* (or *Sehi*), from an inscription which was found in the archeological site of Iruña-Veleia (Gorrochategui 2009a: 548). This inscription has been lost, though, and its reading and interpretation is extremely doubtful. However, two further person names found in the northeast of Alava, *Luntbel*... and *Lutbelscottio*, appear to have the Basque-Aquitanian elements *bel(tz)* 'black' and *-ko* (dim. suffix).

names, such as *Ambatus* or *Reburrus*, have parallels in the epigraphy of the west of the Iberian Peninsula (particularly among the Lusitanians and Vetones). There is also a theonym, *Helass-*, which appears to be associated with the *Selats-* of Navarre, although a change *s-* > *h-* is unparalleled in Basque. In Biscay, most of the names found are Latin (ibid.: 52). Only a few are pre-Roman, but not Basque-Aquitanian.[45]

If we assume that the Caristii and Varduli were totally or partially Basque-speaking in Ptolomy's times, then the absence —or extreme scarcity— of Basque person names must be accounted for by assuming either that, side by side with the Celtic (or Indo-European) population, there lurked a Basque-speaking population which for some reason did not manage or did not feel the need to chisel their person names in inscriptions, or that that Basque-speaking population is actually the one which appears in inscriptions. In this case, they would have celticized their person names without giving up their language and their divinity names. The former option —which does not exclude the coexistence with the latter— fits better into the state of affairs in Biscay,[46] whereas the latter could explain the facts in Alava. As was mentioned above, person names are easily subject to fashions, and the substitution of native names for foreign ones from one generation to the following is copiously attested in Roman epigraphy, as shown by several cases of the Bronze of Ascoli.[47]

By contrast, if we assume that the Caristii and the Varduli, or most of them, were not Basque-speaking, then we are obliged to postulate a migration of Basque speakers into these territories during the Late Antiquity or Early Middle Ages, along the lines of the one proposed by Abaitua & Unzueta (2011). According to these authors, such a migration, to be dated from ca. 500AD on, would be the result of a population overgrowth of a Basque-speaking population

45 In Guipuscoa, only two inscriptions have been found (ibid.: 52). One of them contains the (probably) Aquitanian name *Belteso-*, but it has to be pointed out that this was found near Oiartzun, at the eastern end of Guipuscoa but within the territory of the Vascones in Antiquity, and not far from Aquitania. The other is in Zegama and only contains Latin names.

46 Gorrochategui (1995a: 52) puts it in the following way (referring to Biscay): "As we concluded above for the southwestern corner of Aquitania, the scarcity of epigraphy together with a high number of Latin names in the few Roman settlements indicates a high degree of indigenousness, which in our case is most likely Basque".

47 An often cited parallel is what happens in some regions of Peru, where in some Quechua-speaking areas the person and family names engraved on tombstones are Spanish. This can be summarized under the following formula: "Whereas the cementery speaks Spanish, the market speaks Quechua".

around Pamplona-Iruñea.[48] It would have occurred first westwards, following the axis Pamplona-Iruñea → Vitoria-Gasteiz along the old Roman road Iter XXXIV. Then, after occupying the whole Alavese Plain, that population would have later spread in different directions, including southwards into the High Rioja and northwards into the territories of Biscay and Guipuscoa. Thus, another perpendicular axis would have arisen from south to north or *vice versa*, forming a sociocultural two-way route along which the cult of San Millán would have spread from its original site in La Rioja to the north, leaving a large amount of hagionyms of this saint scattered throughout concrete areas of Alava, Biscay and Guipuscoa.

Such a migration process lacks, at least so far, conclusive archeological support. Abaitua & Unzueta (2011) have argued that a number of archeological sites found in Navarre, Alava and Biscay —the most important of which is that of Aldaieta (northern Alava)— provide support for such a migration process (cf. Azkarate 1999, 2004, Iriarte 1998). These sites are dated to ca. the 6th-7th centuries, and in them elements of Franco-Aquitanian typology have been found. Here, the debate becomes highly interdisciplinary. As is typical with archeological findings, the debate focuses on whether a common typology of tools, dress, weaponry, etc., necessarily implies a movement of populations, or it can be explained by simply assuming a cultural spread.[49]

A potential problem which such an account runs into has been pointed out by Gorrochategui (2002). A late entry of Basque speakers into Alava would be problematic for the explanation of some toponyms, corresponding to early loan words from Latin, which are already attested in the Reja de San Millán (*CSM*, 1025), such as *Guircu* (< Lat. *circu(m)* 'fence'), and *Lucu* (< Lat. *lūcus* '(sacred) wood'), with Basque phonetic treatment. As Gorrochategui admits, an early entry into Basque of a loan word does not warrant the autochthonous nature of a toponym from the period in which it was borrowed by the language, but a clearer case in this respect could be the G. place name *Guetaria* (Gorrochategui

48 As a matter of fact, and as it was pointed out, these authors propose that the incursion started from the Aquitanian territory.

49 The human groups of Aldaieta cannot be simply the evidence of a military incursion, as has been suggested by Böhme (2002), since the analysis of those remains has revealed that they are far from being overwhelmingly masculine, and in fact there are also at least 15 children. Besides, thanks to the analysis of their mitochondrial DNA, we know that many of them were biological relatives, so they appear to make up a familiar group (apud Azkarate 2004: 45-48). Another question is whether they were native of that place or not. According to more recent analyses of the mitochondrial DNA (apud Gorrochategui 2007-08: 1199), certain very specific haplotypes show that they were autochthonous individuals from the north of the Iberian Peninsula.

2009: 550), most probably from Lat. *cētāria* 'fish factory'. This is a town on the Guipuscoan coast, although there is also a *Guétary* on the Labourdin coast, where, apparently, remains of a fish factory dating from the 1st century AD were found, next to a Latin inscription (apud ibid.). It would also be problematic to find an explanation for the suffixes *-aka* and *-ika*, which appear in plenty of Biscayan toponyms and which are of Celtic origin.[50] Had the entry of Basque into Biscay been late, then we should expect to find **-*aga* and **-*iga*, with intervocalic voicing of the voiceless stop, as is usual in Romance and occurs e.g. in the name of *Luzaga* (town in Guadalajara), which apparently corresponds to the *Lutiaca* attested in Antiquity and contains the same suffix *-aka* (cf. Abaitua & Unzueta 2011: 20-21, for a reply to these arguments).

In a nutshell, this is the current state of research. In any case, it seems improbable that Common Basque —a language with few dialectal differences, by definition— was spoken throughout such a vast territory as that from the Nervión to the Garonne, or from the western border of the Vascones to the Garonne. Those who plead the case for a westward geographic expansion after Antiquity will explain the dialectal diversity described in §2.2.2 precisely by means of this process, whereas those who defend that there is no evidence for such a process will explain that diversity as the result of the spread of a *koiné* throughout a territory which was already Basque-Aquitanian or Bascoid, much along the lines of Michelena's (1987 [1981a]) account (cf. §2.3.3).

2.5. Conclusion

In the preceding sections, I have analyzed the intelligibility between Basque dialects, as well as the geographic and linguistic position of each one, taking Bonaparte's fourth classification as a starting point. It should be born in mind that this classification is arbitrary, as in fact any dialectal classification typically is. It has been employed here simply because it could provide us with a systematic procedure to make a tour over the whole dialectal map. In fact, Bascologic studies after Bonaparte have not systematically stuck to the terms used here, especially in subdialectal divisions. Even at the dialectal level High Navarrese and Low Navarrese are often considered as single dialects, Alavese as a dialect independent from Biscayan, etc. Hence, it should not be expected that the terminology employed in the rest of the volume adheres to the terms introduced here. The

50 At least for the suffix *-ika*, Salaberri-Zaratiegi (2011a) has defended a Latin origin. For what concerns us here (why its voiceless stop does not undergo intervocalic voicing), however, its origin is irrelevant.

term *variety*, for instance, will most of the time have a more general sense than the one employed here.

In Section §2.2.3, the different possible ways of analyzing these dialects have been considered, and finally in Section §2.4 the question as to their origin has been summarized, including the problem of how the origin of Basque dialects can be linked to the problems of the languages in the Basque Country and its surroundings in Antiquity and Late Antiquity. As has been shown, both questions are open. The question concerning the earliest splits of CB awaits further discussion on which phonetic-phonological and grammatical features may have played a crucial role in the earliest phases of the dialectal fragmentation, whereas the question of where Basque(-Aquitanian) was spoken in Antiquity will probably become more interdisciplinary.

3. External History. Sources for historical research

Koldo Ulibarri
(University of the Basque Country, UPV/EHU)

3.1. Introduction*

> "The history of a language in the real sense of the word, of Basque as of any other, can be solidly built only upon documents which have been dated and located with the best possible precision […]; the rest —including the reconstructing work to which many, myself the first, have devoted ourselves with a perhaps excessive zeal— is no more than an extrapolation, valid only if it has the firm support of trustworthy data and is useful to explain these data diachronically" (Michelena 1988 [1971c]: 149) [translation by the editor, MMA].

The fact that Basque is a genetically isolated language, and that its literature does not begin until the 16th century, caused Saussure (1916) and Meillet (1925) to assert that it was impossible to know anything about the history of Basque (apud Michelena 1988 [1957a]: 166). However, as Luis Michelena (1988 [1957a]: 168) himself reminds us, Basque does have a history, based on a number of sources which, although not as extensive and rich in their range as we would wish, cast some light on the language even in its old and unattested phases. Unfortunately, this linguistic material has not always been made known, edited and researched in a manner proportional to its philological and linguistic interest.

Historical linguists know that, when reconstructing a language, its oldest sources are essential tools which enable the reconstruction to be built upon a

* This work would not be the same without the comments on the text made to me by Joseba Lakarra. I am also indebted to Blanca Urgell, who was always ready to help and advise on both content and style. It goes without saying that I am thankful to both for their advice and improvement of the original text. I also want to thank Mikel Martínez-Areta for the translation of the text. Mistakes or misconceptions are my sole responsability. Finally, this work has been carried out thanks to a Grant for the Training of Researchers funded by the Basque Government, as well as to the following Research Projects: 1) "Historia de la Lengua Vasca y Lingüística Histórico-Comparada" (HLMV-LHC), Ref. GIC 10/83, IT 486-10, funded by the Basque Government; 2) "Monumenta Linguae Vasconum (III-IV)" [= FFI2008-04516, FFI2012-37696], funded by the Ministry of Education and Science of the Spanish Government; and 3) The Training and Research Unit "Hizkuntzalaritza Teorikoa eta Diakronikoa: Gramatika Unibertsala, Hizkuntza Indoeuroparrak eta Euskara" (UFI 11/14), funded by the University of the Basque Country (UPV/EHU).

solid base, and not just on sheer speculation (see the quotation above). Two possible types of old sources can be found: primary and secondary. Primary sources are texts written in the language under research, and they provide us with information about what the language was like at that particular period. Secondary sources are texts written in other languages, containing words or sentences in the language under research. If there is no old source available at all, the language itself can provide us with some means to know what the forms were like in the past: this is the so-called *internal reconstruction*.

Although these types of sources will be commented on in more depth below, I shall first outline the content of this paper with a short general introduction. The history of the Basque language officially begins in 1545, with the publication of Bernart Dechepare's *Linguae Vasconum Primitiae*, the first book to be written and printed in this language. Nevertheless, the prehistory of the language has two long earlier phases, with no written texts but with other kinds of evidence. The earlier phase belongs to Antiquity; the later one is medieval.

The first data referring to the Basque language date back to the 2nd and 3rd centuries AD, and are a series of god or divinity names, and person names, engraved in Latin on gravestones and votive altars. Along with these, we find some Basque-like place names in Greek-Roman classical sources (§3.2.1). Although these sources do not tell us anything about the grammar and internal structure of the language at that period, they do give us some clues as to its phonetics and lexicon, as well as about the geographical extension of the language in Roman times (an extension which was actually much larger than that of historical times).

After that, we have a hiatus of about seven centuries in which no attestations are found, and we have to wait until the 10th century to find the first "texts" in the strict sense: a couple of glosses in Basque from the Emilian Glosses, which have been dated to around the 10th or 11th century. In fact, some Basque toponymic and anthropoymic attestations are already to be found in the previous century (§3.2). Although the first printed book has been dated to 1545, from the 14th and 15th centuries onwards we already have texts of some importance, in which we find more place and person names, as well as some scattered Basque words in manuscripts or printed documents in Latin and Romance (§3.4.1).[1] From 1545 onwards, the number of attestations increases considerably, although it is still not comparable to that of neighboring languages. In the peninsular Basque

1 Many of these old writings are brilliantly compiled and edited by Michelena and Sarasola in their respective works *TAV* and *ContTAV* (see Michelena 1990 [1964a] and Sarasola 1990 [1981]). Hence, when citing concrete attestations, in many cases I shall do so by means of these works. For the findings corresponding to 1980-1990, see the "Epílogo (1989)" by J. A. Lakarra for the re-edition of both works.

Country, we have barely 500 pages by 1745, and about 30 times as many in the continental Basque Country (see §3.4.2 and ff.). In some cases, thanks to the documents of the period, we know of the existence of some other texts in Basque —whether printed or handwritten— of different natures, which for various reasons have not survived until now.

As for the quality of the texts, it is important to note that the corpus in Basque from the 16th century onwards is —in comparison with neighboring languages— not only late to appear and reduced in size, but also homogeneous in its content and dialectologically asymmetric. Until 1876, 90% of the corpus was of the religious-catechistic genre, and the different Basque dialects are very variously attested over time. Furthermore, despite the scarcity of data in the corpus and its literary and linguistic interest, many texts —in particular long texts from the earliest centuries— have not been made known, philologically scrutinized, and edited until recently, if at all. The spelling system used to transcribe these texts is also a source of problems. Although, in general terms, Basque phonetics presents no major problems in adapting to the Latin alphabet (here the complex system of Basque sibilants are the major exception), we still lack specific research on the spelling systems which were employed over different periods in the attested history of Basque. Existing work focuses on particular (and nearly always relatively modern) periods and dialects, or on particular writings or writers. The only work on spelling which covers a relatively global scope is that of Mujika (2002), which studies the spelling used by continental authors in the 16th-17th centuries. But we have no analogous work for peninsular authors.[2]

In addition to texts and primary sources, there are other types of evidence which can also help us to reconstruct old phases of the Basque language. The study of loan words from both Latin and Romance languages can give us important data about the phonetics and grammar of the oldest phases of the language (§3.2.2). Similarly, comparison between Basque dialects can give us some hints as to the features of the language in a period pre-dating dialectal fragmen-

2 The scarcity of Basque texts in the medieval period poses problems for the study of spelling. However, it would be equally interesting to analyze the spelling methods used to transcribe Basque names and words in Latin and Romance texts (and likewise, the medieval phonetics of those languages), since these incipient phases may help us to understand both the phonetics of Basque and the evolution of its spelling system in different areas.

Such an analysis of the spelling is important if we consider the cultural and geopolitical situation of Basque. It is a language which has never had a single cultural or political epicenter, and the different regions in which it has been spoken in historical periods have belonged to different *scripta* or spelling traditions. This has undoubtedly had some influence on the representation of Basque sounds.

tation (§3.5). Finally, place names are preserved in the regions where Basque is or was spoken in the past. Although much of the systematic collection and linguistic analysis of place names is still to be carried out, in recent years a great amount of work has been done in this field (§3.6).

In the course of this chapter, then, I shall describe the materials available for the reconstruction of the history of Basque, as outlined in this introduction. I shall do so in chronological order, going from the oldest to the most recent sources.[3] As I have said, both the quantity and quality of the texts increases as we move towards the contemporary period. Along with the description of each kind of source, I shall give some examples of what each one can show us.

3.2. The Roman period. Aquitanian or archaic Basque and linguistic contact

The first sources which, at the end of the 19th century, were identified as belonging to a language related to historical Basque, were found in a region which in Caesar's time was called Aquitania, in southwestern France. This is why their language was called the Aquitanian language. In the 20th century, from the 1960s onwards, a further number of proper names were found to the south of the Pyrenees, which can be related to Aquitanian in terms of onomastic and linguistic features. Some of these were found in Navarre and Alava, while, more recently, inscriptions have been found in a territory as distant from the Basque-speaking area as the basins of the rivers Cidacos and Linares (Soria). When analyzing the relationship between Aquitanian and the current Basque language, we have two possibilities: to consider Aquitanian to be an old phase of Basque, or to think of it as a sister language of archaic Basque (the mother of all Basque historical dialects), from which Common Basque would descend. Since evidence is extremely scarce, it is impossible to prove either theory. Hence, from a methodological point of view, the whole corpus —i.e. both the continental testimonies (Aquitanian in the strict sense) and the peninsular ones— are conceived as one single language, which here I shall call, following tradition, the *Aquitanian language*.

A second group of linguistic data can help us to know more about the system of the Basque language in Roman times, and to glimpse some features of

3 The history of the language will be split into different phases, according to the periodization proposed in Gorrochategui & Lakarra (2001: 429). There is as yet no universal consensus on this or any other periodization, but I think that this is the one which best reflects the changes of the language, as well as being the only one based on both internal and external data.

the immediately previous one. These are the Latin loan words from the imperial period. The abundance of these loan words, together with the knowledge of the changes which they have undergone since their introduction into the language, have significantly helped to create the (currently most advanced) reconstruction of the phonological system of Proto-Basque.

3.2.1. Epigraphic evidence

The pioneering work by Luchaire (1877, 1879), and Bascologists such as Lafon (1999 [1956a]) and Michelena (1985 [1954b]), in which the Aquitanian onomastic material was first classified and researched, provided the basis for Joaquín Gorrochategui's (1984) work, in which he gathered all of these materials together and exhaustively analyzed them. Since then, he has commented on every new piece of linguistic material discovered on either side of the Pyrenees (Gorrochategui 1995b, 2009a), both analyzing it and contextualizing it in the framework of what is already known about Old Basque through written evidence or reconstruction. Even if the material discovered in recent years has not drastically changed the linguistic picture of the language known in 1984, it has certainly increased our knowledge of the linguistic situation in the Antiquity.

The sources from the Roman period can be split into two groups. The first group corresponds to a few names of towns and ethnic groups transmitted by classical sources. Although they are of pre-Roman origin, the majority of these can be explained by means of Indo-European languages of Celtic origin (see Gorrochategui 2002: 107). The second group of sources is a scant number of *corpora* with divinity and person names which appear in the Latin texts on gravestones in burial sites and on votive altars from the imperial period (2nd and 3rd centuries AD). In his 1984 work, Gorrochategui gathered over 600 onomastic items, and since then around 100 more have appeared, some from outside the Basque-speaking region. Typical examples of lexical items which appear on these inscriptions as proper names, and which correspond to common nouns in historical Basque, are the following: *Nesca-to* (*neska* 'girl', *-to* 'dim. suffix'), *Sembe-* (*seme* 'son'), *Cison* (*gizon* 'man'), *Seni-cco* (*sehi* 'servant', ~ west. *seiñ* 'child', *-ko* 'dim. suffix'), *Andere-* (*andre* 'woman'), *Bihox-* (*bihotz* 'heart'), all of which were found in the high basin of the Garonne River; *Vmme-sahar* (*ume* 'child', *zahar* 'old'), in Navarre; and *Sesen-co* (*zezen* 'bull'), in Soria. However, we have not yet found any direct evidence (i.e. texts containing sentences with some grammar) of the Old Basque or Aquitanian language (cf. Campbell 2011).[4]

4 Some alleged graphites written on ostraca were "found" in 2005-2006 in the archeological site of Iruña-Veleia (Alava), dated to around the 2nd-5th centuries. These

The geographical distribution of these materials is surprising: few have been found in historically Basque-speaking regions. In Navarre and Alava, we have barely a handful of indigenous names, while in Guipuscoa only one name can be associated with Vasconic features (*Beltesonis*), and in Biscay not even one. Instead, the majority of inscriptions have been found to the north of the Pyrenees, along the valley of the Garonne River, with a high density to be found in the high basin of this river, and on the plain of Gers. By contrast, to the west of these areas, i.e. in Bigorre, Bearn and the continental Basque Country, there are few inscriptions to be found. To the south of the Pyrenees, there are also several names of the Vasconic type in the highlands of Soria (Martínez-Sáenz de Jubera & González-Perujo 1998) and a few in La Rioja (Gómez-Pantoja & Alfaro 2001).

This scarcity of onomastic material poses a number of problems when we analyze the elements, since, as Gorrochategui (1984: 355) remarks, very often there is more than one possibility of segmentation. Likewise, the absence of any contemporary related language to which the data can be compared (as is possible between Celtiberian and Gaulish) makes their analysis even more difficult. As a result, after segmenting and analyzing the elements, we often find ourselves in a *vacuum* of evidence and linguistic parallels, unless we resort to historical Basque (ibid.: 356). Moreover, since these names have been transmitted in Latin epigraphy, they are adapted to the linguistic parameters of Latin, with Latin noun inflection (Gorrochategui 1993: 610).

Even though these proper nouns are secondary materials, they do give us some clues about the language, especially as to its phonetics and phonology.

---

were initially presented as a turning point in our knowledge of Old Basque, for they were texts written in everday language in both Latin and Basque (along with some texts in hieroglyphic Coptic!). However, due to a number of paleographic, iconographic and linguistic inconsistencies, these texts have been proved to be unequivocally false.

From the linguistic point of view, some features present in the graphites are clearly inconsistent with what we know about the Basque of that period, thanks to both data allegedly contemporary with the ostraca and to later data. For instance, some words or constructions found on the ostraca are not attested in Basque until the 18th, 19th or 20th centuries, and so are not present in 16th or 17th century texts. The same holds for the graphites in Latin. But it is even more striking that the texts do not provide any new features for Basque diachronic linguistics. In other words, the linguistic data of the Iruña-Veleia evidence do not allow us to make any progress in our knowledge of the proto-history of Basque. We find more *hapax*, archaic constructions, etc., in 16th century texts than on those ostraca (see Gorrochategui 2009a: 539-540). Furthermore, the type of texts written on the ostraca is extremely rare and incoherent with its alleged function, and in most cases is unparalleled in the Roman world. Gorrochategui (2011) takes an overall look at the problems posed by the ostraca of Iruña-Veleia.

Thus, they provide us with real proof of certain features which can be glimpsed in the internal reconstruction of the historical data. These include features associated with the phonological inventory and its restrictions (two series of sibilants, the existence of *h*, the absence of initial *r*-, the contrastive opposition between *-n-* and *-nn-*, etc.), phonetic changes (*-nb-* > *-m-*, *th-* > *h-*, etc.), and so on. In terms of morphological features, the evidence provided by these onomastic features is restricted to derivational suffixing. As for syntax, they simply tell us about the order of elements within NPs (noun + adjective). However, lexical analysis is more fruitful, since a comparison with historical Basque allows us to interpret and group the elements into clear semantic fields, such as kinship terms, animals, colors, numerals, etc. It also helps us to understand the onomastic system of the imperial period, as well as its continuation in the Basque-speaking region during the Middle Ages. Likewise, it provides us with data for reconstructing Basque words and their elements, which would otherwise remain unexplained, as in the case of *ume* 'child' < ***(b)on-be*** and *seme* 'son' < ***sen-be***.

3.2.2. Latin loan words

Another potential source of information for prehistoric Basque is the analysis of how loan words are borrowed. We know that Basque was in contact with other Indo-European languages before the arrival of Latin, but few Indo-European pre-Latin loan words have been preserved, if any (see Gorrochategui 1987). To explain this fact, Michelena (1988 [1972a]: 123) suggests that Latin loan words simply substituted the pre-Latin ones, a tendency borne out by the structural proximity between the two languages and their two thousand years of constant contact. Therefore, for any promising analysis of this point we must stick to the loan words which Basque has borrowed from Latin from the very beginning of their contact (around the 1st century BC), along with later loan words from the neighboring Romance languages. At present, the study of loan words is not as advanced as would be desirable, and although loan words have been one of the key pieces of evidence for reconstructing Proto-Basque phonetics, we have as yet no monographic work about them. Among those studies which have analyzed the Latin element in the Basque language, we have Michelena (1988 [1972a], 1987 [1974]), Knörr (1995), Echenique (1997, 2008) and Lakarra (2012c).

Although we cannot always distinguish Latin loan words from Romance ones, the study of loan words is very helpful for research into Proto-Basque. The relatively detailed knowledge that we have about Latin and the Romance languages and their evolution over time enables us to estimate when a given loan

word was borrowed, and to consider what that implies for the study of the Basque phonological system. In the oldest loan words, the Latin timbre is kept, without any of the further changes undergone by the Romance languages. Thus, in those words we find a velar pronunciation of *g* and *c* before *e*, *i*, with no palatalization, as in Romance: *bake* 'peace' < Lat. *pāce(m)*, *gela* 'room' < Lat. *cella*... Similarly, Latin loan words which have undergone the stop initial voicing characteristic of Old Basque can also be considered old (cf. the examples cited above). And those which show the result of the lenition processes *-n-* > *-Ø-* (*balea* 'whale' < Lat. *ballēna*), *-nn-* > *-n-* (*anoa* 'daily ration of food' < Lat. *annōna*), *-l-* > *-r-* (*a(i)ngeru* 'angel' < Lat. *angelus*), and *-ll-* > *-l-* (*gaztelu* 'castle' < Lat. *castellu(m)*), which took place in Basque some time in the Middle Ages (see Michelena 1988 [1964b: 36 ff.]) were obviously borrowed before such changes occurred.

The study of these Latin words and their subsequent evolution in Basque helped Martinet (1974 [1950]) and Michelena (1988 [1957a], *FHV*) in their reconstruction of the phonological system of old Basque. Nevertheless, as yet there has been no systematic attempt to date loan words and, by means of the well-known evolution of Romance, to establish a general relative chronology of the phonetic changes in prehistorical Basque. Guiter (1989), however, represents an attempt in that direction.

3.3. Medieval Basque. Centuries of lights and shadows

After the fall of the Roman Empire in the 5th century, there is a period of around three or four centuries, usually called the "dark" centuries, in which written production drops. We have to wait until the 8th-9th centuries to see how the so-called *Carolingian Renaissance*, especially in the Church and its monasteries, rekindles the writing tradition in Europe, giving rise to new written evidence of the Basque language, always in Latin texts.

In the earliest documentation of the Middle Ages, the quality of these sources is not very different from that found in the Roman period. Again, they are mainly place and person names, the number of which increases over the centuries. These materials slowly begin to appear at the time in which monasteries, most of them associated to the incipient Kingdom of Pamplona, are founded, and begin to take over lands in the Basque Country. The first territories which include anthroponyms and toponyms with Basque features are La Rioja, Castile and Alava, all three of which were partially Basque-speaking regions in that period, with monasteries such as Valpuesta, San Millán de la Cogolla, Albelda, etc.; Navarre, with monasteries such as Leire, Irantzu, La Oliva, Iratxe, etc.; and

Huesca, with the monastery of San Juan de la Peña. This is followed by Biscay, and later by Guipuscoa. The medieval documentation of the continental Basque Country is not as old nor as abundant as that of the peninsular one, and, furthermore, it has been studied to a lesser extent (the cartulary of Saint-Jean de Sorde, the censi of the 14th and 15th centuries, etc.).[5]

The person names which can be found in this documentation are to some extent a continuation of the onomastics of the Aquitanian inscriptions. However, place names already present some dialectal features which we later find in historical Basque, with lexical variants and archaisms (such as the maintenance of aspiration, probably general in some old period) which separate the peninsular western dialects (Biscayan and Alavese) from the eastern ones. As was the case with the person names from the Roman period, the information provided by these data is mainly phonetic, due to the intrinsic features of onomastic material. However, in a few cases we can already find the first attestations of noun declension features (such as the well-known *Lope Jaun Ortire semea* 'Lope, son of Sir Orti' (Iratxe, Navarre, 1125), where *-re* is an archaic form of the genitive), as well as common nouns used as nicknames: *Andia* 'big', *Anaie* 'brother (of a boy)', *Arçaia* 'shepherd', etc. (see Michelena 1987 [1969]: 92 ff.). The detailed study of medieval place names has also helped us to understand the grammaticalization process of the Basque definite article *-a* from the 3rd grade demonstrative **har* (Manterola 2008), and the formation of the plural (Michelena 1987 [1971b]).

Although all of these sources are fragmentary and usually consist of simple lexical items, in a few lucky cases we find sentences, or at least something close to them. The first attestations of this nature are two short sentences in the Emilian Glosses, dated to the 10th or 11th century (de Bustos 2004: 304). These are *jzioqui dugu* and *guec ajutuezdugu*, which even today pose problems of interpre-

5 There have been attempts to compile all of the Basque materials of this period, scattered in the cartularies of the different monasteries of the Basque Country and its surroundings. Apart from Michelena (1990 [1964a]; = *TAV*), we can mention Arzamendi (1985), with references to older compilations, and Líbano-Zumalacárregui (1995-2000). However, there is still no work which analyzes and sequences the phonetic, morphological, syntactic and dialectological data of these materials, thus composing a fully-fledged monograph of Medieval Basque (see, however, the various works by Irigoyen 1985-1995, as well as Orpustan 1999 and Salaberri-Zaratiegi's 1999 review of this). Another pending task is the updating of the lists of material, whereby new data can be discovered when the already known (but chaotic) material has been put in order and scrutinized, as Peterson (2007) has done for the Monastery of San Millán de la Cogolla. An attempt to geographically arrange and systematize all of this material has been undertaken by Reguero-Ugarte (2012).

tation (*TAV*: 41). Dating from the 13th century, some Basque words and expressions also appear in the General Privilege of Navarre (*TAV*: 51).

Finally, another medieval source which is worth mentioning is the list of words compiled by the French pilgrim Aimery Picaud, who, in the 12th century, asked some words in Basque on his way to Santiago de Compostela, and wrote them down, glossed in Latin (today in the *Codex Calixtinus*; *TAV*: 49). Similar lists, usually compiled by foreign travellers, proliferate in the Renaissance. In Basque philology, we have those made by the German Arnold von Harff (15th century; *TAV*: 63) and the Sicilian Lucius Marineus Siculus (first half of the 16th century; *TAV*: 146).[6]

3.4. The Modern Era. Texts in Basque

Although literature in Basque appears quite late (no lengthy texts are to be found until the 15th century, and no printed work until the middle of the 16th century), it has been well established that, before this, there existed a rich oral literature in Basque (ballads, dirges, sayings, etc.), even if it has no written attestations until the 16th century (Garibay, Ibargüen-Cachopín, etc.). This genre has partially survived up until now. Due to the nature of the transmission, and as is usual in this type of literature, these texts often preserve archaisms from earlier periods (see Lakarra *et alii* (eds.) 1983, Lakarra 1983, 2008b).

Although the texts in Basque are much richer and more abundant in the Modern Era than in earlier periods, they still pose problems. One such problem is that the literature is still more limited than that of neighboring ones. Another problem is the dialectal asymmetry to be found in the corpus. The different Basque dialects are very variously attested in terms of both the period of their first appearance and their preservation over time. In fact, there is no geographically complete corpus of texts until the 19th century, when for example the Alavese dialect was nearly extinct and Southern High Navarrese languished. In the case of Low Navarrese, it is well attested in the 16th century (thanks to Bernart Dechepare's work), but we do not find it documented again until around 200 years later. Until 1745, the continental literature prevails over the peninsular one, but after that the peninsular one takes the lead, at least in terms of the number of texts. On the other hand, some dialects, even those which are as important

6 Although we usually differentiate medieval from Archaic Basque (which will be discussed in the next section), it is the nature of the registered material, rather than the state of the language itself, which motivates a division between these two phases. Both the morphology and the syntax of the Emilian Glosses are very similar to those of Archaic Basque.

from the diachronic point of view (see §3.4.3 and §3.5) as the Pyrenean ones, are not significantly attested until the 19th century.

Another drawback of these texts is that, until practically the 20th century, they are almost exclusively of a religious nature, often limited to catechisms and prayer books. This considerably limits the variety of lexical fields, as well as, for example, the sectors of the verbal system, such as the potential or the subjunctive, which can be found in a given text (for statistics on Basque texts across periods and dialects, see Sarasola 1976: 179 ff., who categorizes them into translations, original works and reeditions). In any case, new texts often appear in archives or libraries, either among unclassified manuscripts, or in old books whose content remains unknown to Basque linguists.

In the following sections, I outline what can be found in Basque texts and the motivations for their writing. I do not list them exhaustively (see the index of authors, texts and sources at the beginning of this volume, in Abbreviations), but I briefly describe the type of text and its main features in each case, as well as the geographic-dialectal deficiencies of each period (for further information on this, see Urquizu 2000 and Villasante 1979 [1961], the most complete available histories of Basque literature).

3.4.1. 15th century – 1600. Archaic Basque

It is not until the 15th century that we have some relatively extensive texts, which provide us with the features of Basque from that period in some detail. These texts were found in the documentation of the archive of the Kingdom of Navarre, and are the following: a letter which can be dated, according to the first editor of the text, to 1415 (*ContTAV*: 75); and a popular prayer written down before 1425, of which several popular versions still survive today (*TAV*: 57). Already in the 16th century, we have a letter written in Biscayan Basque from Mexico by the bishop Fray Juan de Zumarraga to his sister in the Basque Country (1537, *ContTAV*: 97), and some other correspondence relating to espionage activities in the Hispanic-French war (1595-98). In the course of the same century, we find prolific historiographic studies by authors from the western territories, which typically focus on the historical events of the 15th century in Biscay, Alava and Guipuscoa. Among such authors, we can include Esteban de Garibay, Ibargüen-Cachopín, Francisco de Mendieta and Lope Martínez de Isasti.[7] In

7 Already in the 15th century, we find the Biscayan Lope Garcia de Salazar as a forerunner of Basque historiography. To these authors whose works have —in a more or less complete form— survived until now, we should add a number of earlier works of which we know only the title and/or the author's name, thanks to citations to these found in the pre-

these chronicles, written in Romance (Early Modern Castilian), a number of glossed words and etymologies, sentences and especially old songs in Basque are gathered. Most of these are written in Biscayan (and some in the linguistically close Alavese dialect), since both the authors and the narrated facts frequently belonged to areas where this dialect was spoken. Basque was typically employed in such texts to reinforce the narration or to reach a stylistic climax (cf. *TAV*: 66 ff.).

From around the end of this period, we can find the first attempts to prove that Basque was the first language spoken in Spain. Thus, by means of glossed etymologies of place names from across the Iberian Peninsula, authors like Andrés de Poza (1587) and Baltasar de Echave (1607) provide us with further evidence of the language at this stage.

A key factor which contributed to the Basque language being finally written down, as with many other languages in Europe, was the spread of Renaissance ideas, which had already had an influence on Bernart Dechepare (Arkotxa & Oyharçabal 2009). The Renaissance was a period in which vernacular languages gained ground against classical languages, and in which scholars turned their attention to the former. The peculiar nature of Basque soon sparked off curiosity and interest in the rest of Europe. To the previously mentioned sources of von Harff and Marineus Siculus, here we should add a larger enterprise undertaken by Niccolò Landucci in 1562: the *Dictionarium Linguae Cantabricae.* This is an early Spanish-Basque dictionary, which was, until the discovery of Lazarraga's manuscript in 2004, the main source of evidence for the 16th century Alavese dialect. Two of the informants used by Landucci to write the dictionary were apparently from Vitoria-Gasteiz, so the dictionary presumably reflects this variety. The Basque language was also included —usually with texts from Leiçarraga's *New Testament*— in the typical plurilingual collections of the Renaissance (Knörr 1996: 136).

Apart from the language itself, oral tradition and folklore were equally appreciated as a source of knowledge. At the end of the 16th and the beginning of the 17th century, collections of sayings and proverbs were gathered in several regions of the Basque Country: Biscay (Garibay, ca. 1592, and the *Refranes y Sentencias*, 1596), Guipuscoa (Martínez de Isasti, ca. 1625; however, this is HN from a dialectal point of view), Soule (Jacques de Bela and Bertrand de Sauguis, in the first half of the 17th century, and Arnaud d'Oihenart, 1657), etc. It should be stressed that both old poems and proverbs like these are products of an oral tradition, and hence they often preserve, like fossils, many features of the lan-

served texts. Some of these are Bedia Zirarruista, Aguayo Gebara, Peña Galdocha, Alfonso Castilla, Gonzalo Arredondo, etc. (Arriolabengoa 2008: 22).

guage from earlier periods. That is why, although many of these sources were not written down until the 17th century or even later, they also provide some evidence of the Basque spoken in the 15th and 16th centuries.

Along with the incipient concern for vernacular languages, the increase in literacy in the Renaissance also contributed to the increase in written literary production in Basque. We have short poetic excerpts in Basque, both handwritten and edited, in texts written in other languages (see *TAV* and *ContTAV*). But, from 1545, we can speak of a Basque literature in the real meaning of the word, for in that year a book in Basque was printed for the first time. This was a collection of sixteen poems written in Eastern Low Navarrese by Bernart Dechepare. Between 1567 and 1605, the writer Juan Perez de Lazarraga, from Alava, penned —but did not publish— a bucolic novel and a number of poems on Renaissance topics. The discovery of the manuscript in 2004 (albeit damaged, with some excerpts missing) was a milestone in Basque philological studies, from both a literary point of view (because of its stylistic singularity) and a linguistic one, since it has enormously increased our knowledge of the 16th century Alavese dialect (Lakarra 2004a, Bilbao *et alii* 2010).

Another factor which exerted a great influence over Basque was the ecclesiastical conflict triggered by the Reformation and the subsequent Counter-Reformation in 16th century Europe. In what concerns us here, the consequence of that conflict was that vernacular languages took on an unprecedented importance in pastoral labor. Thanks to the endeavors of the Reformation to translate the sacred texts into vernacular languages, we have the works of the Calvinist Joannes Leiçarraga, who, under the patronage of the Queen of Navarre Jeanne d'Albret, translated the *New Testament* (1571) into Basque, along with a didactic and catechistic work. The linguistic relevance of this work is that, as he explains in the prologue, he attempts to write "in a language which should be comprehensible to everyone, not in the language of a particular place" [translation by the editor, MMA]. In the peninsular Basque Country, Catholics —spurred by the Counter-Reformation— also began to use Basque in their catechistic labors, and from this period on we know of two doctrines: one by the Navarrese Sancho de Elso (1561, today lost), and one by the Alavese Juan Perez de Betolaza (1596, still preserved).

To sum up, in this period of Archaic Basque, the bulk of the corpus consists of sources which cover the western part of the Basque Country (both Biscay and Alava), as well as, further to the east (but west of the continental Basque Country), Low Navarre and Labourd. In the center (Navarre and Guipuscoa) and, at

the eastern end of the continental Basque Country, in Soule, we find a gap in the evidence which is only partially made up for by a few micro-texts.[8]

3.4.2. 1600 – 1745. Old Classical Basque

Leaving aside the (especially morphological and syntactic) changes in the language itself (Gorrochategui & Lakarra 2001: 429), the 17th century also brings change in the quality of available sources. The Catholic Counter-Reformation was triumphing in the Basque Country and also, from the moment that the French king Henry IV left Protestantism for Catholicism (1593), in France. Joannes Leiçarraga's attempt to create a Basque *koiné*, based on continental Basque and probably conceived for that region, did not get any further, and the tendency followed by later authors was for each author to use his own local variety. This tendency was particularly widespread in the peninsula. In the continent, in spite of the failure to form a fully-fledged *koiné*, authors tended to assimilate their language to the coastal variety. From a linguistic perspective, then, the texts from 1600 onwards offer us a wider geographical scope. Stylistically and thematically, however, these texts are still overwhelmingly of a predominantly religious nature.

In this period, important texts appear in hitherto unattested, or barely attested dialects, including catechisms and mass-handbooks in Southern High Navarrese (Juan de Beriayn, 1621, 1626, Francisco de Eliçalde, 1735), and in Souletin (Athanase de Belapeyre, 1696). In this latter dialect, however, there already exist earlier texts (some letters written between Souletins and Roncalese dating from 1615-17, and a small book by Arnaud François de Maytie, 1676), as well as those which we shall see below (Arnaud d'Oihenart, 1657, Jean de Tartas, 1672). We have to wait until the beginning of the 18th century to find the first lengthy texts in Guipuscoan; namely a doctrine written in the variety of Ordizia (Joseph Ochoa de Arin, 1713), followed by another one written in the variety of Hernialde (Juan de Irazusta, 1735).[9] Likewise, the production of doctrines continues in earlier attested dialects, such as in Biscayan (the anonymous *Viva Je-*

8 In the case of Guipuscoa, the translation of the Psalm *Miserere* and some other poems (*ContTAV*: 77). In Soule, we have the three previously mentioned collections of sayings, although their writing actually belongs to the 17th century.

9 There are, however, letters from Azkoitia (Satrústegui 1987: 33), and Azpeitia (*ContTAV*: 124), corresponding to the end of the 16th and the beginning of the 17th centuries, respectively, and an extremely interesting text from Azkoitia, to be dated to 1614, containing several ordinances and exercises which were to be practiced by the nuns of a local congregation (cf. Urquizu 2008-09).

sus, ca. 1640 [southwestern Biscay], Martin Ochoa de Capanaga, 1656, Nicolás Zubia, 1691, and Martin de Arzadun, 1731 [Duranguesado, or the Durango area]); and in Labourdin (Esteve Materre, 1613, and Sylvain Pouvreau, 1656).

As far as the language is concerned, the literature which proliferated in the continental Basque Country during this period is of the utmost relevance, especially that which is written in Labourdin. The quantity of works written from 1617 to 1643 greatly exceeds that of all earlier periods, and their stylistic quality increases.[10] These authors use Leiçarraga's work as a referent, and they appear to mutually influence each other (although this second point has not been sufficiently studied; cf. Salaberri-Muñoa 2000). They write in a diatopically rather homogenous variety (that of the towns of Donibane Lohizune and Sara), since they believed that each author should write in the variety of his home town (but, as expected, in a language stylistically higher than that of ordinary people).[11] We have previously mentioned the doctrines of Materre and Pouvreau, the first writers in Basque who were *euskaldun berriak*, i.e. who did not speak Basque as their mother-tongue. To these doctrines, which include prayer books in verse like those by Joannes Etcheverry de Ciboure (1627, 1630, 1636), Bernard Gasteluçar (1641), Joannes Haranburu (1635), etc., we can add translations of texts in vogue (Christobal Harizmendi, 1660, Jean d'Aranbillaga, 1684), or works of their own creation, which reach their literary zenith with Axular's *Guero* (1643).

Apart from these religious works, there also appear secular texts (in Labourdin, too) about navigation (Piarres d'Etcheverry, 1677) and veterinary science (Mongongo Dassança, 1692), although the latter was not printed. The influence of the group affected the whole of the continental Basque Country, and indeed the authors of the Souletin literature which was emerging at that time (Oihenart, 1657, Tartas, 1666, 1672) explicitly state the influence of Labourdin in their works. They also have some influence in the peninsula, as in the previously mentioned texts by the Navarrese Beriayn. But it is Larramendi, in the following

10 This group of writers was long considered to be a literary school called *School of Sara*, for it was the variety of this town in which these authors chose to write their work. Although nowadays most Bascologists, for historical and textual reasons, do not consider this group of authors as a "school" (Urgell 1991, and Salaberri-Muñoa 2002), it is still possible to find this name in the literature.

11 Despite this, it seems that during this period the Basque of that region had some prestige, as attested by Voltoire in the following statement: "…comme ordinairement de trois ou quatre lieues en distance de Terre ou Pais, il ce void & ce treuue changement de langage, presque par touts endroids, & quil est impossible de comprendre ou rapporter dung particulier vng general, Tant y à que pour le Basque ie lay traduict tel quil se parle au melheur endroidt des Basques, qu'est à S. Iehan de Lus & Siboure, que te sera pour aduys…" (Voltoire, ca.1620).

period, who introduces this literature —including a spelling system which owes much to it— to the peninsular Basque Country, and who recognizes its value, at least as far as Leiçarraga and Axular are concerned.[12] This recognition has held in learned literature up until now.

Although the peninsular literature is not comparable in either quality or quantity to the Labourdin one, some texts do appear in the peninsula during this period, apart from the doctrinal ones. In Biscayan we find literature, mostly handwritten, which includes love motifs in verse (already documented in the archaic period; *ContTAV*: 87), and historical texts. Some of these even reach the printing press, such as the *Egiaren Kanta* ('Eguia's Chant', ca. 1638), which is apparently the first evidence of cordel literature in Basque (cf. Urgell 1986).[13] There is also religious poetry, with texts written mainly in Alavese and High Navarrese (*TAV*: 111), some of which were even published far away from the Basque Country (in Seville, cf. Ulibarri 2011), along with the production of Christmas texts, such as songs for the common people (Pedro de Ardanaz, 1674), and the bilingual Christmas carol by Simón Martínez de Ochoa (1692). This trend will continue in the following period. We also have the first theater piece (Pedro Ignacio de Barrutia, from the beginning of the 18th century).

As we have seen, I have divided the documentation from this period into two groups: the continental and the peninsular. Although this division is mainly due to dialectal reasons, and to the differences in quantity and, above all, quality, of the sources found in each region, these are not the only differences. Since the two literatures were written in different cultural spheres (the peninsular literature in a Castilian one, and the continental literature in a French one), the authors in each of the traditions devised different spelling systems based on the orthography of their neighboring Romance language. Very often, the authors show inconsistency in this respect, or conceive spelling proposals which have little success. We have to wait until the following period to see the first attempt —by Larramendi— to create a unified graphic system for both sides of the frontier.

An emerging text type of this period is that of books for language teaching. Due to their nature, these provide very interesting data for the history of Basque. At the time, trade was widespread in Europe and Latin was losing its prowess as

12 The prologue of his *Diccionario Trilingüe* can be considered to be the first history of Basque literature, as in this dictionary Larramendi provides data about all Basque books known by him (which are practically the whole body of literature up to his time).

13 The *Bertso Bizkaitarrak* ("Biscayan Verses", 1688; cf. Lakarra 1984b) probably belong to the same genre, although we cannot be sure that they were ever printed. There are more cases of verses transmitted in a handwritten format, often kept in manuscripts as evidence in slander trials.

lingua franca. Learning languages became important, and hence learning methods and small dictionaries proliferated throughout Europe. In the Basque Country, during this period we can find language handbooks in several dialects (although only some in Labourdin were actually printed): in Labourdin (Voltoire, 1620), in the Biscayan of Bilbao (Rafael de Micoleta, 1653), and in the Guipuscoan of Donostia-San Sebastián (Jossef Domingo de Lubieta, 1728). There are also more elaborate grammars, such as the handwritten ones in Labourdin by Pierre d'Urthe (ca. 1714), and Joannes Etcheverry de Sara (ca. 1712); the printed ones by Martín de Harriet (1741; in L), and Larramendi (1729; in G); as well as the lost grammars written by Dominique Bidégaray (ca. 1675; in LN), Bela (1667; in S), and Melchor de Santa Inés Oyanguren (1715; in G?). Along with grammars, lexicographic works also began to appear, with dictionaries like the one by Pouvreau (1665), or the lost one written by Etcheverry de Ciboure. Some others exist as parts of previously mentioned grammars: Harriet (edited with his grammar), d'Urthe (ca. 1715), and the lost ones written by Bela, Bidégaray, Etcheverry de Sara (1712), and Oyanguren.

The proliferation of grammars and dictionaries in this period must be understood in the context of the literature of that time, i.e. as necessary tools for literary production. And yet fully-fledged patronage for these types of texts was lacking, probably due to the lack of potential purchasers. Because of this, most of these works were not published at their time, and unfortunately many were later lost.

3.4.3. 1745 – 1887. Early Modern Basque

1745 is a milestone in the history of the Basque language, because in that year Manuel de Larramendi printed his *Diccionario Trilingüe* (Spanish, Basque and Latin). By then, he had already written an apology of the Basque language (1728), and a grammar (previously mentioned, 1729). As a consequence of this dictionary and literary approach, it is in this period that the peninsular literature made the most progress in terms of its style, number of authors and edited works. By contrast, the Labourdin literature, so prolific in the previous century, experienced a significant decline. This was already perceptible by 1643 (Oyharçabal 2001: 14), but became more pronounced during this period because of the boost received by French in its replacement of Latin as the language of culture. The literature was not to recover until around 1840.

As in the previous century, 18th century Basque literature mainly consisted of religious texts. Doctrines in already-known varieties of the language continued to be translated or edited (i.e. in B, G, HN, S, L), and the Low Navarrese

dialect, not attested since Dechepare (1545), re-emerged with the didactic text *Kadet eta Bettiriño* (post 1745) and the translation of the work by A. Rodríguez (López, 1782; although this Low Navarrese is dialectally very close to Souletin).

The peninsular literature, which took the earlier classics of the continent as a model, was the emerging literature in terms of both quantity and quality. It was more prolific in sermons and religious texts, written in both prose and verse. Its greatest exponents were the Jesuit missionaries, gathered around Larramendi, and the Franciscans of the missionary college of Zarautz, but its exponents also included secular priests. The sermons produced during this time make up a large corpus, especially in Guipuscoan and Biscayan, but were not published at the time, and, more recently, have only been published to a small extent. Likewise, in this period we also find texts in Alavese, which either deal with religious topics (J. García de Albeniz, 1778), or are secular poems (J. B. Gamiz, 1760).

After the publication of his *Diccionario Trilingüe*, Larramendi, along with the Jesuits Agustín de Cardaberaz and Sebastián de Mendiburu —who were inspired by his linguistic ideas— created a literary language (*literary Guipuscoan*) based on the different Guipuscoan varieties, since each author had kept the dialectal differences of his own variety. The Labourdin literature of the previous century had a considerable influence upon this language. As a result, this literary Guipuscoan evolved and gained in importance throughout the 19th century. It had a modern spelling system (again, based on the continental system of the previous century) and its creators took special care with its lexicon. They tended to reject loan words and to favor patrimonial words, sometimes inventing neologisms. As a result, this Guipuscoan dialect had a great influence upon all peninsular literature, and became the most prestigious dialect (Zuazo 1988: 146), used by authors as important as Juan Bautista Aguirre de Asteasu or Juan Ignacio de Iztueta.

Literature in the Biscayan dialect, in turn, was founded by Juan Antonio Moguel and Pedro Antonio de Añibarro around the turn of the 19th century. It reached its literary peak in the first two decades of the 19th century (1802-1816), with authors such as Vicenta Moguel and Fray Bartolomé. These Biscayan authors even had some influence, in terms of lexicon and some other features, on later Guipuscoan authors (e.g. Iztueta). Likewise, Liçarraga de Elcano began to write in the High Navarrese dialect, which until then had only been attested in doctrines.

One of the characteristics of these authors is that several of them were actually able to write in more than one dialect. Thus, Cardaberaz wrote in both Guipuscoan and Biscayan, and Añibarro in Biscayan, Guipuscoan and High Navarrese. This is a consequence of the main function of Basque literature in this period: to prepare preachers, for whom these sermonaries were written, to teach

the doctrine and faith to common people in their language. There are other outstanding works, however, which do not share that aim, such as the *Euscararen berri onac* by Cardaberaz (the first book of style in Basque), the *Peru Abarca* by Moguel (sometimes considered to be the first novel in Basque, although it was actually conceived as a didactic work in dialogue), the *Guipuzcoaco dantza gogoangarrien condaira* (a treaty on Basque dances) and the *Guipuzcoaco provinciaren condaira* (a history of Guipuscoa) by Juan Ignacio de Iztueta, the *Egunare eusquerascoa* (a Basque calendar) by José Paulo Ulibarri, etc.

Meanwhile, in the continent, in addition to the previously mentioned translation by López and the dialogue *Kadet eta Bettiriño*, we have examples of non-catechistic texts from this period in both Souletin (philosophical writings by Jusef Eguiategui, 1785), and Labourdin (the didactic text in dialogue *Kadet eta Bettiriño*, post 1745, the didactic text *Birjina edo donceil christaba*, 1783-84, and Martin Duhalde's *Meditacioneac gei premiatsuenen gainean*, 1809). All of these texts were already influenced by Larramendi's dictionary. Also from this period, we have the first examples of conserved *pastorals*, a type of popular theater in the Souletin dialect, which have still only been partially edited, but which, due to their linguistic archaisms, are important from a diachronic perspective.

Literary production in the continent, however, was renewed under the prominent figure of Joseph-Augustin Chaho. He initiated this literature with journalistic works in Basque published in his journal *L'Ariel*, and was followed by several authors, most of them Low Navarrese (Jean Martin Hiribarren, Duvoisin, Francisco Laphitz, Gratien Adema "Zaldubi"…). As a result, the literary Basque of the continent took on Low Navarrese dialectal features, forming a language sometimes labeled *Navarrese-Labourdin*.

With the ideas of the Enlightenment, increasing importance was given to the education of children. As a result, during this period didactic texts like fables (La Fontaine, Esopo…) began to be translated into Basque on both sides of the Pyrenees. This is the case with the texts translated into Biscayan by Juan Antonio Moguel, his niece Vicenta Moguel (1804) and Mateo de Zavala (ca. 1840), and with the translations into Guipuscoan by Agustín Iturriaga (1848). In the continental Basque Country, we have the translations by the Souletin Jean-Baptiste Archu (1848) and the Labourdin Leonce Goyetche (1852).

Through Wilhelm von Humboldt's works, the European linguistic elite began to take an interest in the Basque language, and from this period we have the first philological works carried out by both foreign (Julien Vinson, Edward S. Dodgson…) and native (Julio de Urquijo, Fidel Fita…) authors. Vinson's bibliographic (*Essay d'une Bibliographie de la Langue Basque*, 1891) and philological work is especially remarkable, while the other scholars' work is also out-

standing. Dodgson and Fita were engaged in editing old manuscripts (Micoleta, Joseph de Jesús María Araquistain), and re-editing books (Capanaga), while Urquijo rescued multiple texts, and provided them to linguists, in both critical editions (*Refranes y Sentencias*, the work by Etcheverry de Sara…) and facsimiles (several classical books in the *Revista Internacional de Estudios Vascos*).

The European interest in the Basque language also leaves its trace on the figure of Prince Louis-Lucien Bonaparte, nephew of Napoleon I. He studied Basque dialectology, and, contacting mostly local priests from multiple regions, asked them to translate Bible texts or other materials. He also gathered old books and manuscripts, and compiled a large number of texts in all varieties of the language. It is thanks to him that we have the first substantial attestations of the Pyrenean dialects (Aescoan, Salazarese and Roncalese),[14] which, due to their archaic nature, are extremely important for historians of the language. We also owe to him texts written in other varieties which were poorly attested until his lifetime. However, although the number of texts enormously increases in diatopic terms with Bonaparte's texts, religious themes are still dominant. This is because the texts which the prince entrusted to his collaborators for translation (J. A. Uriarte, J. P. Duvoisin, M. Intchauspe..., in fact also known in the general literature of the period) were usually Christian doctrines or Biblical texts.

Grammars and dictionaries were necessary to satisfy both the needs of childhood education and the interest of European linguists. Some were published in the 19th century, while others remained handwritten. Those written by Larramendi (his grammar in 1729 and his dictionary in 1745) were already difficult to find by the beginning of the 19th century. Moreover, the grammar was only for the Guipuscoan dialect, and thus was of little help to writers of other dialects, and further did not fit into the new ideas about grammar. Towards the end of the 19th century, four dialects were considered the "literary dialects", and we can find grammars for all four of these throughout the century: Biscayan (Añibarro, ca. 1804, Zavala, 1848), Guipuscoan (Luis Astigarraga, 1825, Francisco Ignacio Lardizabal, 1856, Juan María de Eguren, 1867), Labourdin (Fleury Lécluse, 1826, Francisco Jáuregui, 1850), and Souletin (Archu, 1848, Luis Gèze, 1873). Two more grammars which were later added to these include the one by Arturo Campión (1884), which collects the four dialects together, and the comparative one by Willem Jan van Eys (1879). In the case of dictionaries, leaving aside those which appear as supplements to the grammars, they are basically an attempt to modernize and correct that of Larramendi, although many of them remain in manuscript form. Examples of such dictionaries include Jose Francisco Aizkibel (1885), van Eys (1873), the etymological one by Pedro Novia de

14 Regarding the latter dialect, we now have the aforementioned letters from 1615-17.

Salcedo (1887), etc., and the more recently published ones by Añibarro, Iztueta, José María Odriozola, Maurice Harriet (this last one unfortunately remains unedited).

Due to the influence of J. A. Moguel in the peninsula and Chaho in the continent (in turn influenced by European Romanticism), Basque traditional literature began to attract interest. Thanks to this increase in interest, songs and oral texts began to be collected (Chaho, 1844, Francisque Michel, 1857, J. D. J. Salaberri, 1870, Iztueta, 1826, José Manterola, 1877-80), and the *bertsolaris* or *koblakaris* became popular.[15] These authors wrote verses which people learnt by heart and repeated, or sold them in fairs in the form of *bertso-papera* (similar to the cordel literature or *littérature du colportage* of the neighboring languages). During this time, the names of *bertsolaris* and bards such as Pierre Topet Etchahun, José María de Iparraguirre, Pedro Mari Otaño, G. J. I. Bizkarrondo "Bilintx", Jean-Baptiste Elissamburu, etc., started to become well-known.[16]

The total loss of the Basque *Fueros* in 1876 —already considerably reduced in 1839— raised interest in the Basque language as a means of maintaining the identity of the Basque people. This brought about an increase in texts, thanks to which Basque literature was modernized. There was poetry, drama, novels, but also journalistic texts, which contributed to a diversification of the data in this period of the language. It is this rich thematic variety of texts which causes Larramendi's neologisms to proliferate during this period.

In terms of orthography, this period sees the first attempts to create a single spelling system for both sides of the frontier. This task had already been undertaken by Larramendi. His spelling system was first imitated by his Order fellows, and later by the rest of the authors. But these orthographic rules changed over the 19th century, especially among Low Navarrese authors in the continent (who imitate Chaho's way, e.g. using <z> and not <ç>, <ñ> and not <gn>; see Zabaltza 2011: 38), and in the peninsula, above all in Biscayan literature. Biscayan changes inherit the views of Astarloa, and are reflected in the second edition of Añibarro's *Escu-liburua* ('Handbook', 1821; the first one is from 1802), e.g. using <k>, and not <qu>, <y> and not <j>. These changes were gradually assimilated by Guipuscoan authors such as Iztueta and Lardizabal (Urgell 1987). Most Guipuscoan authors, however, kept Larramendi's spelling system, e.g.

15 These are a kind of bard who improvises verses about a given topic, usually informal or from everyday life. They already existed in earlier periods, according to news-reporting songs of the 15th and 16th centuries and the remains of literature in verse available to us from the 17th and 18th centuries, such as the oral literature gathered from the period described in this section, of unknown author.

16 Many of the works of these authors can be found in the collection *Auspoa*, created by Antonio Zavala.

Serafín Baroja, "Bilintx", Marcelino Soroa, Victoriano Iraola... All of these modifications were systematized by José Ignacio Arana in his project of a Basque spelling system (1872, 1890), which was the point of departure for the unifying enterprise of the following period.

3.4.4. 1887 – 1968. Late Modern Basque

The foundation of the journal *Eskualduna* in the continent is the starting point of this period. In fact, the literature of the early 20th century was marked by the press written in Basque (whose forerunner was Chaho, 1848). Journals were written not only in the continent (*Eskualduna*, *Gure Herria*...) and in the peninsula (*Euskalzale*, *Ibaizabal*, *Euzkadi*, *Argia*...), but also in the overseas Basque community (*Eskualdun Gazeta*, 1885, *California-ko Eskual Herria*, 1893-1898). In the continent, the flourishing period which had begun in the previous phase (Villasante 1979 [1961]: 157 ff.) continued until the Second World War, with authors who followed in the path of the earlier ones, such as Jean-Baptiste Hiriart-Urruty, Jean Barbier and Jean Etchepare.

At the turn of the century, Basquism was undergoing a change in mentality which, along with other effects, stimulated a yearning for the primitive language, i.e. a desire to recover it at its most perfect stage, with no corruption from other languages. These ideas bubbled up especially in Biscay, where proposals were made for the promotion of a completely regular grammar (which, however, did not succeed in society), and attempts were made to banish Latin and Romance loan words, regardless of their antiquity in the language, and to substitute them with neologisms. A forerunner of the creation and promotion of a regular grammar was Resurrección Mª de Azkue (*Euskal Izkindea*, 1891), although he soon gave it up, and also Manuel Arriandiaga (Laka 1987). This type of language has been called *hyper-Biscayan*, because it was based on that dialect, fostered its most distinguishing features, and had a great influence upon the Biscayan writers of the following decades: Raimundo Olabide, Evaristo Bustinza "Kirikiño", Esteban Urkiaga "Lauaxeta", etc.

As for the efforts to create a lexicon free from foreign influences, two trends were soon formed. One was led by Azkue himself, whose aim was to use the Basque word compounding and word derivational rules, to create new words. The other was led by Sabino de Arana y Goiri, the founder of Basque Nationalism. He created, and encouraged others to create, neologisms according to particular rules which were devised by him, and which actually made the words unintelligible. Neologisms, especially those of the second trend, were more widely accepted than the 'totally regular grammar' project. Some of these suc-

cessful neologisms were created to express modern concepts, while others were intended to replace loan words, some of which had been totally assimilated by Basque in previous centuries. This diffusion, however, was largely based on political rather than linguistic reasons.

Azkue and his followers also aimed to eliminate loan words, but, although they also —unsuccessfully— tried to create neologisms, their major effort was to gather the words used by the common people, and re-use them for both modern and traditional concepts. This group of scholars (Juan Bautista Eguzkitza, Gerhard Bähr, Nikolas Ormaetxea "Orixe"...) has left us a large number of works which are extremely important for the general history of Basque, including dictionaries, grammars, traditional literature, and dialectal collections, which constitute a solid base for any diachronic research. Azkue himself was author of several works in all of these fields, such as the *Cancionero popular vasco* (1915), *Morfología Vasca* (1923-25), and *Evolución de la lengua vasca* (1935). He was also promoter of the *Triple Cuestionario* (1935). But it is his *Diccionario Vasco-Español-Francés* [*DVEF*] (1905) that brought about a paradigm shift in both Basque lexicography and the history of the Basque lexicon. The impact of his work on the lexicon used by Basque authors can easily be observed by comparing certain novels written before his dictionary (*Auñemendiko lorea*, 1898) to those written subsequently (*Kresala*, 1906, *Garoa*, 1912).

During this period, the Basque provincial councils founded *Eusko Ikaskuntza* (1918), an institute created to promote Basque culture. One of its first challenges was the unification of the Basque language and its orthography. In order to undertake this task, *Euskaltzaindia* (the Royal Academy of the Basque language) was created from *Eusko Ikaskuntza*, one year later. Its motto was "to research and maintain the language" (*Iker eta jagon sailak*). As for orthography, here we should mention that there had already been one attempt to unify it, in the *Congreso Ortográfico de Hendaya* (1901), an event which gathered writers and intellectuals from both sides of the frontier. Their starting point was the proposal made by José Ignacio Arana (see §3.4.3), but no consensus was reached as to a single orthographic system, due to discrepancies which were more political than linguistic. *Euskaltzaindia* initially carried out projects for the compilation of materials which would be useful for the unification of the language (the collection of lexical and dialectal information, the study of old texts, grammar projects), but the Spanish Civil War (1936-39) prevented the completion of these projects.

*3.4.5. 1968 – Today. The unified Basque language (*euskara batua*)*

After the Civil War, we have to wait until the 1950s to see the re-emergence of literary production in Basque and a renewal in studies about the language. During this decade, the unifying project reappeared, so that finally, in 1968 —on the 50th anniversary of the Royal Academy—, a congress was held in Arantzazu (Guipuscoa). In that congress, the foundations were laid for a common literary language known as unified Basque or *euskara batua*, starting from the guidelines of Luis Michelena's programmatic paper. Michelena himself, José Luis Álvarez-Enparantza "Txillardegi", Gabriel Aresti, Luis Villasante and Jon Mirande, among others, were the fathers of this unified language. They wrote papers on it, and produced literary works which promoted its divulgation.[17] This language enabled Basque to be more efficiently learned by the non-Basque speaking population of the Basque Country, and paved the way for the production of a literature common to the whole of the Basque territory. Since then, the *euskara batua* has been in a gradual but consistent process of development, beginning with the orthography (1968) and the verb (1973, revised in 1997), and continuing with a grammar (1987), the unified dictionary (1995, by revising the lists from 1968, 1974, etc.), pronunciation (1998), etc. The first monolingual dictionary —leaving aside the unfinished one by G. Aresti— is the *Hauta Lanerako Hiztegia* by Ibon Sarasola (1984b, re-edited in 2005 as the *Euskal Hiztegia*). Apart from its work on unification, *Euskaltzaindia* continues to gather interesting materials for the historical research of Basque, with works such as *El Atlas de las Hablas Populares* (1983) and the *Diccionario General Vasco* (1987-2005; = *OEH*, cf. Euskaltzaindia 1987-2005).

Basque has now become fully integrated into society, thanks to, among other factors, the following: the Statute of Autonomy and the creation of the Basque Government (1979) (undoubtedly the main driving force of the unified language); the Basquization of the Administration; the creation of bilingual programs in the compulsory education system and later in university education (along with linguistic research on Basque); and the existence of mass media in Basque, in both public and private spheres (and in both *batua* and specific dialects). Likewise, the progress in new technologies has enabled this language to open itself to the world, with a rather high percentage of use on the Internet, in proportion to the number of its speakers.

17 The influence which Gabriel Aresti had upon the unification is worth highlighting, as Michelena used to say that the first work ever written in *euskara batua* was his poem *Maldan behera* (= 'Downhill') (Villasante 1994: 86). This poem was presented at a literary competition in 1959, nine years before the Arantzazu Congress and the official birth of *batua*.

3.5. Basque dialectology

Within the diachronic research of a language, dialectology deserves a special place, as Sever Pop pointed out in the foreword of his book on this topic:

> "Le linguiste Antoine Meillet a affirmé avec raison qu'il n'y a pas d'histoire de la langue sans une dialectologie et surtout sans une géographie linguistique complète et bien établie (Bull, de la Soc. de Ling. de Paris, t. XXX, 1929, p. 200), [...] Leur importance capitale pour toute étude linguistique est aujourd'hui indéniable; ils constituent des témoignages vivants qui ont pour le linguiste au moins la même valeur que [...] les documents d'archives pour l'historien" (Pop 1950: IX).

As we have seen, in the case of Basque, dialects are a very important part of language research, for, since the language had no common literary language until the end of the 20th century, old texts are written in the different dialects. Since Basque dialectology is described in more depth in another chapter, here we describe only a few of its features.

Azkue split Basque into seven dialects. To these, we should add Alavese, which at the time of Azkue, and indeed of Bonaparte, was about to disappear. However, during his time, the testimonies in Alavese from previous centuries, particularly those of Landucci and Lazarraga, were not yet known to scholars.

To reconstruct Old Basque by means of the comparison between Basque dialects has a major drawback: the differences between dialects are not very big for a linguist. In fact, in terms of comparison they are desperately small, and hence reconstruction cannot progress as much as we would wish (Michelena 1988 [1964b]: 3). This is due to the fact that the mother language from which all dialects branched off is not as old as might be expected by those who consider Basque dialects to be already formed by the Roman period (so that Ptolemy's Caristii would correspond to historical Biscayan, the Varduli to historical Guipuscoan, and the Vascones to historical Navarrese) (cf. §2.3.1). Since Michelena (1987 [1981a]: 54), it is usually assumed that the common language must be dated to some point around the 5th-7th centuries, but the lack of sources for this period makes it impossible to set a more specific date.

Moreover, we should add two problems of using dialectal comparison as a source for reconstruction, both of which should have been made clear from the preceding sections: the late attestation of Basque dialects, and the asymmetry of their documentation (see §3.4). In spite of these problems, comparison between Basque dialects does provide data for the reconstruction of Old Basque. As is the case with any language with a number of dialects, diatopic diversity is a spectrum whereby some varieties preserve features which other varieties have lost, or are losing. This is especially, but not exclusively, the case of peripheral —hence archaic— dialects. In some cases, we can apply Bartoli's principle,

whereby if a feature is present at both geographical extremes of a language, then it is old (§2.2.2.10). Thus, if we compare Biscayan to Roncalese or Souletin, we sometimes find convergent features (both lexical and grammatical) which undoubtedly belonged to Old Basque. A lexical example is B. *i(ra)tzartu* 'to wake up' and *iratzarri* in Souletin, but *esnatu* in central dialects. A grammatical one is the 3rd p. pl. erg. verbal marker, very roughly and skipping details *-e* in Biscayan (and adjacent varieties) and Souletin (and adjacent varieties) but *-te* in central dialects.

The gathering of dialectal data, first spurred by Bonaparte in the 19th century (the materials have recently been edited by Pagola *et alii* 1992-99), has continued throughout the 20th century. Two outstanding works are Pedro Yrizar's *Morfología del Verbo Auxiliar Vasco* (1991-2008), and the recent linguistic Atlas of Basque under publication by *Euskaltzaindia*, as well as a great array of other works, on a smaller scale but equally interesting. As a matter of fact, even today the gathering of dialectal material provides data of great interest for the history of the Basque language and its evolution, especially in marginal areas. The research methodology of historical dialectology is also breaking through in Basque research (Ulibarri 2010). Thus, the study of old texts and the comparison between them allows us to describe the evolution and spreading of a given linguistic innovation in the language. We can also draw a picture of the different linguistic choices made by each dialect in each period. Thus, texts of obscure origin, which by extralinguistic criteria alone cannot be assigned to a particular variety, can be located on the map with a degree of certainty. For example, the language of the catechism known as *Viva Jesus* (dating from the 17th century, with no geographical origin which can be ascribed to it on extra-linguistic grounds) has been located between the regions known as South Uribe and Ayala, to the south of Bilbao (Ulibarri 2010).

3.6. Toponymy

It is a common procedure to use place names as a means to explore the past of a language. Every researcher of a language will resort, to some extent, to onomastic materials in order to reconstruct its past, but this method is particularly important when direct written evidence is scarce. This is the case with languages which have left no written record, but whose place names have partially been recorded in neighboring languages. Although this kind of source is potentially of the utmost importance, such data must be handled with care, for they are attested in a language which is alien to the language under study, and this gives rise to

certain problems associated with its phonetics and transliteration.[18] This is the case with some cities of Antiquity, such as *Oiasso* (in eastern Guipuscoa, but belonging to the Vascones at that time), or *Calagurris* (in Lower Rioja, considered as a city of the Vascones at least by Strabo and Ptolemy), both of them attested in several forms depending on the author or the manuscript.

In the case of languages which are historically closer to us, and which have a more ample range of testimonies, toponymy can be contrasted with historical data, allowing us to interpret its lexical, phonetic, and sometimes even grammatical, elements. These data can help us to know about the oldest stages of the language (§3.3), but also about earlier phases, especially in areas where the language was lost. Moreover, place names have the advantage that they can be located to a particular place, so they guarantee that the linguistic data found in that place name really existed in the speech of that place. As is well known, toponymy tends to be more conservative than everyday language, so that we must take care when dating the appearance or absence of a particular feature. For instance, we find forms like *Galdacano* or *Ochandiano* in texts, but the forms traditionally used by Basque speakers have been *Galdakao* and *Otxandio*, with the Basque treatment of intervocalic *-n-* (> *-Ø-*). Furthermore, toponymy does not cover every aspect of language, but only parts of the lexicon and phonetics, and, with luck, some glimpse of the grammar (as in the place name *Euri-dacar* '(lit.) (it) brings rain'; Kontrasta, Alava, 1540).

In the study of Basque, it should not be forgotten that it is onomastic testimonies which first appear in the written history, although these are found in texts written in another language (Latin or Romance). This is not as strange as it might seem at first sight, as in languages with a written tradition richer than ours we find a similar state of affairs. The fact that toponymy —along with anthroponymy— is one of the first sources of evidence for Basque, means that its study is an extremely important element for reconstructing the earliest phase of the language, that is, the phase which includes Medieval Basque and the configuration of historical dialects.

18 A case in point is Old European, a language usually considered by scholars (since Krahe's work) as belonging to the Indo-European family, but introduced to Europe by a population wave prior to the classical Indo-European groups (Celtic, Italic, Greek, Germanic...), and preserved in river and mountain names all over Europe. In the 1990s, Theo Vennemann came up with the theory that these place names are related to Basque (1994, 2003). However, his theories suffer from serious methodological and theoretical flaws, from both the Basque (Lakarra 1996b, 2012a), Indo-European (Kitson 1996) and Semitic (Baldi & Page 2006) points of view (cf. now the critical essays included in Udolph (ed.) 2012).

The study of place names has also contributed to establishing the geographical extension of the language at different prehistorical phases. Thus, thanks to the studies of the Basque toponymy of La Rioja by Merino-Urrutia (1978 [1962]) and studies of the Pyrenees by Corominas (1960, as well as his works *Onomasticon Cataloniae* and *Estudis de Toponimia Catalana*), we know that there existed Basque-speaking spots in areas which historically have not belonged to the Basque Country, but for which Basque was extra-linguistically reported by some historical source. Merino-Urrutia's work has enabled us to establish the variety of Basque spoken in the region (dialectally western), but unfortunately there has been neither a comprehensive continuation nor a revision of the methodology, despite the fact that Basque linguistics has made considerable progress in recent years (cf., however, Salaberri-Zaratiegi 2011b).

By contrast, considerable progress has been made in the study of the place names of areas where Basque is still spoken or has been spoken until recently. Here, for obvious reasons, there exist better toponymic data in terms of both quantity and quality, and the corresponding variety once spoken there is also better attested in texts. Michelena also made contributions to this field (1987 [1971b], 1988 [1984]), as did Odón de Apraiz (1920, 1930, and, in collaboration with Knörr, 1978), and Henrike Knörr (1985). In 1994, the first monographic study was carried out on the Basque of a region with no existing written texts (Salaberri-Zaratiegi 1994), and more recently others have been made, such as those on the Roncal Valley and the neighboring Ansó Valley (in Aragón, López-Mugartza 2008); and on the toponymy of Vitoria-Gasteiz (Martínez de Madina 2009-10). Although these studies usually focus on the dialectal features of the language, they are still interesting, and necessary for the diachronic research of Basque, since as we have already seen, dialectology also serves diachrony.

In this field, *Euskaltzaindia* has carried out an important work with its *Onomasticon Vasconicae*, a collection which gathers studies on toponymy from the whole of the Basque Country, including some of those mentioned above. As for the study of continental toponymy, two outstanding researchers are Eugène Goyheneche (his PhD thesis *Onomastique du Nord du Pays Basque*, written in 1966, has been recently published as Goyheneche 2011) and Jean-Baptiste Orpustan (2006).

3.7. Conclusion

Some general considerations should suffice to round off this chapter. As Michelena pointed out (see the opening quotation), the reconstruction of a language

requires a solid foundation of evidence, and only the knowledge and analysis of such evidence will enable us to ensure that this reconstruction is sound.

To sum up the problems posed by the sources for Basque, the main ones are that such texts are scarce in comparison to those of neighboring languages, and that direct testimonies are late. This is especially relevant considering that a language related to Basque —perhaps even to its very earliest phase— is attested near the region, albeit in a fragmentary manner, in the Roman period (2nd-3rd centuries AD). Likewise, although sources began to proliferate during the 15th and 16th centuries, these are both diachronically and geographically asymmetric. Consequently, in the majority of the cases there is no continuity of evidence in the same region, and it is not unusual to find gaps of over one hundred years with no attestation at all for a particular area. As for its geographic extension, until the 19th century we lack a range of sources wide enough to cover all of the dialects at the same period. Finally, old sources are thematically homogenous until 1800, since 90 percent of the corpus is religious. This prevents us from obtaining a complete picture of the language, with regards to its lexicon, verbal morphology and syntax, until the 19th century.

As a result, one of the greatest challenges for those of us who work on sources for the history of Basque is that many of the texts —even those from some of the oldest periods— have still not been edited under the minimal parameters demanded by a "philology of precision", following Meillet's terminology. However, since the pioneering works from the beginning of the 20th century (Urquijo, Vinson, Dodgson...), and especially since Michelena's work, much progress has been made in the editing and analysis of Basque texts. In addition, more texts have been found, and the general production of editions —of both discovered and previously existing texts— has increased in quality and in quantity. Research into secondary materials is also on track to provide interesting results for historical research. As we have discussed, loan words and their consequences for a chronologization of historical facts have not yet been exhaustively studied. Dialectological studies are currently at a fruitful stage, and toponymic data are also being gathered and historically analyzed. In these two realms, the predominating perspective of dialectologists and toponymists is usually dialectological and synchronic, but diachronists can also profit from such research.

4. Phonetics and Phonology

Ander Egurtzegi
(University of the Basque Country, UPV/EHU)

4.1. Introduction*

Fifty years after its original release, *Fonética Histórica Vasca* (= *FHV*, Michelena 1977 [1961]) continues to be the main reference insofar as Basque historical phonology is concerned. Although the seminal work of Michelena is, undoubtedly, the most significant milestone in the history of Basque diachronical phonology (maybe even within Basque diachrony itself), an update of the latest research postdating this work represents indeed an interesting endeavor.

There are, however, different ways to undertake such an investment of time and effort. In this work, I have chosen to go beyond a simple compilation of the articles on Basque historical phonology written in the last fifty years, by producing a new survey. Thus, this chapter is a summary of the most significant work on the history of the Basque language, but it also introduces comments and discussion on it.

The chapter is composed of seven sections including this introduction. Section §4.2 presents the three main stages of the language usually mentioned in the bibliography (namely Old Proto-Basque, Proto-Basque and the modern language) as well as their (reconstructed) phonemic inventories. Section §4.3 discusses the vowels and the processes that have affected them, while Section §4.4 focuses on the consonantal inventory in a similar way. Section §4.5 describes the phonotactics of the language and the possible consonant clusters, and Section §4.6 describes the different kinds of accentuation found in the modern dialects, proposing a possible relative chronology of the accentuation types and

* This work has been supported by the Basque Government by means of the predoctoral grant BFI 2010-385, EJ-GV's consolidated research group "Historia de la Lengua Vasca y Lingüística Histórico-Comparada" (HLMV-LHC), Ref. GIC 10/83, IT 486-10, MINECO's "Monumenta Linguae Vasconum (III)" [FFI2008-04516] and "Monumenta Linguae Vasconum (IV): Textos arcaicos vascos y euskera antiguo" [FFI2012-37696], as well as UPV/EHU's UFI11/14. I would like to express my gratitude to Joseba Lakarra, Juliette Blevins and Gorka Elordieta for their continuous support and useful advice. Likewise, I would like to thank Carlos Cid and José Ignacio Hualde for reading this work and making interesting suggestions. I am truly thankful to all of them. It goes without saying that any errors in this work are entirely mine.

their variants. Eventually, a brief Section §4.7 with some general conclusions is provided.

4.2. The Basque phonemic inventory: now, then, and before then

4.2.1. Modern Basque dialects

Modern Basque dialects vary to some extent in their phonemic inventory. Nevertheless, a more or less common inventory would include the phonemes usually linked to Standard Basque. These segments are presented in Table 4.1 and Table 4.2, together with some other added sounds, which correspond to those found in only a limited number of dialects (or even in one single dialect). The latter are represented in parentheses.[1]

From a typological point of view, an unusual distinction within the Basque consonantal inventory is the phonemic opposition between six different voiceless sibilants: an apical, a laminal and a post-alveolar, along with their affricated counterparts. In Souletin and some Low Navarrese varieties, there is also a series of voiced fricative sibilants —and even an apico-alveolar affricate— which are phonemically distinct (cf. Hualde 1993b).

When carefully studying the segments which are present in only a few varieties, the historical linguist must distinguish between the segments which represent an innovation and those present in an older stage of the language but which are only peripherally present today after having become lost in most of the modern dialects.

1 Hualde (*GB*: 15) offers a simpler table containing only the segments found in Standard Basque. In addition, he includes a table with the phonemic inventory of the easternmost dialect, Souletin (*GB*: 18), usually described as the most deviant variety of the Basque language.

Table 4.1. Consonantal inventory of modern Standard Basque

		stop		fricative		affricate		nasal	lateral	tap	trill	glides
		voice-less	voi-ced	voice-less	voi-ced	voice-less	voi-ced					
bilabial	labial	p (p^h)	b					m				
labio-dental				f								
apico-dental	apical	t (t^h)	d									
apico-alvelar				s̻	(z̻)	ʦ̻	(ʣ̻)	n	l	ɾ	r	
	laminal (alveolar)			s̺	(z̺)	ʦ̺						
palato-alveolar	predorsal		ɟ		(ʒ)							
palatal		c (c^h)	g		(j)			ɲ	ʎ			j
	postdorsal	k (k^h)		x								w
	laryngeal			(h)	(ɦ̃)							

Table 4.2. Vowel inventory of modern Standard Basque

	front		central	back
high	i (ĩ)	(y) (ỹ)		u (ũ)
mid	e (ẽ)			o
low			a (ã)	

4.2.2. Michelenian stages

In fact, as we go back in time, we find a very different scenario: the reconstructed Michelenian Common Basque (cf. Michelena 1987 [1981a]) from the beginning of the Middle Ages (5th-6th centuries AD), and Proto-Basque (cf. *FHV*), dating to around the 3rd century BC. After separating the wheat from the chaff, fewer than half of the segments found today had a phonemic status at those early stages, and most of these only after having suffered significant changes (cf. Martinet 1974 [1950], *FHV*, 1988 [1951b]). As a matter of fact, the PB. system was not grounded on the opposition between voiced and voiceless stops or fricative and affricate sibilants, but on a more general fortis versus lenis distinction which involved almost the entire consonantal inventory (apud *FHV*, although some authors, such as Hualde 1999 or Egurtzegi, in this paper, are skeptical about this reconstruction).

As Table 4.5 shows, the reconstructed vowel inventory contains the same vowels as in modern times. The sixth vowel from Souletin (the front high rounded vowel /y/) as well as the nasalized series are both regarded as innovations, although they have a very different chronology. The consonant inventory in Tables 4.3 and 4.4, on the other hand, looks very different shape in the stages of the language reconstructed by Michelena from the modern one, with an opposition dividing the system into two groups —the aforementioned fortes and lenes— in an almost regular way. However, this is still incomplete given the lack of the labial fortis in the inherited lexicon, as discussed by Michelena (*FHV*: 261, §4.4.2.2).

The first proposal for such an opposition comes from Martinet (1974 [1950]), who hypothesized that Proto-Basque may have had a system analogous to that of the Danish language. He proposed an opposition between two series —namely fortis and lenis— that had different phonetic realizations according to the position of the word in which they were produced. The fortis series was produced as aspirated and voiceless in a prominent position and as plain voiceless preceding an unstressed vowel, while the "soft" or lenis stops were realized as devoiced stops in a prominent word-initial position and as voiced fricatives in an

unstressed position. Table 4.6 illustrates the hypothesized phonetic realization of the stops, according to Martinet (1974 [1950]: 533).

Table 4.3. Consonants involved in the fortis/lenis opposition[2]

	stop			sibilant		sonorant		
	labial	dental	velar	apical	laminal	nasal	lateral	rhotic
fortis	(P)	T	K	t̪s̪	t̻s̻	N	L	R
lenis	p	t	k	s̪	s̻	n	l	r

Table 4.4. Consonants absent from the fortis/lenis opposition

	labial	palatal	post-dorsal	laryngeal
fricative	(f)			h
nasal	(m)			
glide		(j)	(w)	

Table 4.5. Vowels reconstructed for Proto-Basque

	front	central	back
high	i		u
mid	e		o
low		a	

Table 4.6. Hypothesized phonetic realization of stops

Phonemes	Initially (stressed)	Intervocalically (unstressed)
/p/	[p^{h}-]	[-b̥-]
/b/	[b̥-]	[-β̞-]
/t/	[t^{h}-]	[-d̥-]
/d/	[d̥-]	[-ð̞-]
/k/	[k^{h}-]	[-g̊-]
/g/	[g̊-]	[-ɣ̞-]

Taking this system as a starting point, Martinet proposes a word-initial lenitive process involving the fricativization of the aspirated stops. This is followed

2 Here we follow Michelena's notation (*FHV* and elsewhere), using uppercase letters to denote the fortis and lowercase for the lenis series.

by a subsequent debuccalization that could have resulted in the complete loss of the segment —i.e. /t^h-/ > /θ-/ > /h-/ (> Ø)—, along with an initial voicing of the lenis series, which the author attributes to neighboring Romance pressure, since "in Latin pronunciation, the articulatory energy was equitably distributed along the chain" (ibid.: 54 [my translation, AE]). Michelena (i.a. 1988 [1951b], *FHV*) accepts and expands this system with the addition of the sonorant segments to which, interestingly enough, he does not attribute any kind of contextual divergence in their phonetic realizations.

To the consonants in Table 4.6, two different groups of consonants may be added. Firstly, the aspiration, which was most probably part of that stage of the language, even though it did not form part of the opposition proposed by Martinet (1974 [1950]) and expanded by Michelena (1988 [1951b], *FHV*). Secondly, the segments /m/ and /f/ along with the glides *yod* and *wau* and the aforementioned /P/ were not as usual as the other segments and were probably of later introduction. While the former was almost undoubtedly part of the previous stages of the language, the latter labial segments seem to be secondary in older times (some maybe even marginal) and it is possible that they were mostly present in the borrowed lexicon.

Michelena (*FHV*: 374) cites the adaptation of Latin and Romance medial voiced stop clusters as evidence of a tense/lax contrast in opposition to the modern voiced/voiceless contrast. The fact was that all geminates and heterorganic voiced stop groups were adapted to voiceless stops with the place of articulation of the prevocalic segment, as in Lat. *abbas* > *apaiz* 'priest' or Old Sp. *cobdiçia* > *gutizia* 'desire, whim'. An alternative account, however, would be that medial voiced stops were realized as approximants —as they are in the modern language and already were in the 12th century (cf. §4.4.2.1)— and that voiced stop clusters could be perceived as the only existing voiceless stops, rather than word-internal voiced approximants. Along the same lines, he also cites the regular adaptation of Latin /f/ as Basque /b/. However, the latter may have been the only labial consonant available in the language and thus the only suitable option, given the lack of /p/, as proposed by Michelena himself.

4.2.3. Lakarran Old Proto-Basque

If we move into even older times, some centuries before the time estimated by Michelena for PB (approx. 3rd century BC) we encounter a very different state of the Basque language which we can label Old Proto-Basque. This stage has mainly been reconstructed by Lakarra (1995, 2005, 2011b, etc.), and is a direct consequence of his research on the old canonical root structure which he found

to be a monosyllabic CVC, lacking any consonantal clusters or diphthongs (cf. Lakarra 1995 and elsewhere).

Thus, Old Proto-Basque has much simpler phonotactics than the following stages of the language, allowing only two possible consonantal positions (i.e. C_1VC_2) instead of the initial/medial/final distinction made in both Proto-Basque and modern Basque. Thus, we distinguish between onset and coda consonants in Old Proto-Basque, as illustrated in Table 4.7.

Table 4.7. Consonants in Old Proto-Basque according to their syllabic position

	stop			fricative			sonorant		
	lab	dent	vel	apical	laminal	glottal	nasal	lateral	rhotic
onset	b	(t), d	(k), g	s̻	ʂ	h	n	-	-
coda	-	-	-	s̻	ʂ	-	n	l	(r, ɾ)

It is uncertain what the total number of proposed stops and rhotics should be, or what was the nature of the plosive series. The reconstructed monosyllabic roots involving voiceless stops (such as **tor* or **kar*) are scarce, but cannot be derived from other segments either.

Little research has been carried out on the Proto-Basque vocalic system, and thus it is even harder to know the shape of that system in previous stages of the language. Until new results are found, a 5-vowel inventory is reconstructed by default, since it is the one found in the modern stages of the language and in any written document, except for those written in the Souletin dialect, with a front high rounded vowel as well as a nasalized series (§4.3.2 and §4.3.3).

4.3. Vowels

4.3.1. Main vowels

Basque vowels do not seem to have changed to a great extent over time. The five vowels divided into three different degrees of openness are found in the majority of the modern dialects —/a, e, i, o, u/— directly correspond to the vowels reconstructed for the Michelenian Proto-Basque.[3] An uncertain point for this 5-vowel reconstruction was the alternation between /i/ and /u/ and the possibility

3 To this common inventory, certain (mostly local) innovative segments may be added, namely, the Souletin high front labialized /y/, the nasalized series from Souletin and Roncalese, and the Biscayan "sensibly palatal, relaxed and slightly labialized" *a* from Gernika, as described by Navarro-Tomás (1923: 50).

of such an alternation being the consequence of an older phoneme, an option which had already been discarded by Michelena (*FHV*: 73, §4.3.4.1).

Some varieties maintain a vocalic gemination or even a length distinction (cf. Michelena *FHV*: 111-112), which originated from the encounter of two similar vocalic segments in derivation, in composition, or after the loss of an intervening consonant. This distinction is especially frequent in the Biscayan dialect, although not necessarily limited to it. Examples of this include Old B. *errekaak* 'rivers, streams', *domeekaak* (< Lat. *dominica*) 'Sundays', *arabaar* 'Alavese' and B. *haar* > *har* 'worm'. Contraction is the most usual outcome in other dialects.

Vowels are characterized by two seemingly contradictory tendencies (apud *FHV*: 50): they tend to be affected by neighboring sounds but they have not undergone any deep or generalized changes.

Following the observations of Gavel (1920) and Navarro-Tomás (i.a. 1923), Michelena points out that most vocalic permutations occur either between front vowels (i.e. /a, e, i/) or back vowels (/o, u/), although there are some other processes which affect vowels grouped according to their height[4] (*FHV*: 59). Likewise, he notes that the ease for vowel interaction increases in proportion to the number of syllables present in the word (*FHV*: 70).

4.3.2. Nasalized Vowels

The Basque vocalic phonological inventory has historically possessed a series of phonemically contrastive nasalized vowels. These principally resulted from the loss of the nasal environment of the vowels surrounding an intervocalic /n/, which would have become a nasalized aspiration at around the beginning of the Middle Ages (cf. *FHV*: 300-303, Igartua 2011, §4.4.1.3.1 and §4.4.3.1.2).

Today, these nasalized vowels are still found in the modern Souletin dialect exactly as they were found in the defunct Roncalese variety, which disappeared in the 20th century. However, according to Michelena (*FHV*: 48), they represent an "archaic particularity on its way to extinction" even in Souletin, due to their unstable condition. Souletin lacks the nasalized /õ/, since this systematically became /ũ/ after the development of the rounded high front vowel /y/ (*FHV*: 48). Souletin nasalized vowels appear in two different environments: surrounding a nasal (or nasalized) consonant, as in *khatiña* [kʰatĩɲã] 'chain', *gazna* [gaz̪nã] 'cheese', *ama* [ãmã] 'mother' and *ahate* [ãh̃ãte] 'duck'; or in a stressed final position, after the apocope of an old /-n/ (Hualde *GB*: 31): *salu* [s̪a'lũ] (cf. Sp.

4 See, e.g., reciprocal metatheses involving vowels (cf. Egurtzegi 2011, §4.3.4.6)

salón) 'living room', *bedezi* [beð̩e'ʂĩ] (< Fr. *medecin*) 'physician', *ardu* [ar'ð̩ũ] (< **ardano*) 'wine'. Although the former context may be a coarticulatory outcome with no phonemic consequences, the latter clearly contrasts with word-final non-nasalized vowels in other oxytones.

Historically, nasalized vowels are undoubtedly found in Biscayan until at least the 16th century, as documented by various contemporary authors, and they are present in Biscayan texts such as *Refranes y Sentencias* despite the lack of any form of notation to represent nasal phonemes in several words (cf. *FHV*: 48-49). They are also explicitly attested in texts such as those of Madariaga and the manuscript from Garibay.

Michelena assumes that nasalized vowels were present in further varieties of the language (*FHV*: 50), noting that they can be attested as far back as the 12th century in a word list included in *Le guide du pélerin de Saint-Jacques de Compostelle*,[5] in words such as *ardum* 'uīnum' and *arraign* 'piscem'.

As for the loss of this nasalized series in most dialects, this was probably the consequence of a natural internal development: a reinterpretation of those nasalized vowels as a /Vn/ sequence; an example of sound change by Chance (Blevins 2004) or a hypercorrection *à la* Ohala (1993: 250, 255), as in *arrãi*, *artzãi* becoming *arrain* 'fish' and *artzain* 'shepherd' in Biscay (cf. Egurtzegi 2011). Nevertheless, this process may have been conditioned by neighboring languages, given that varieties in contact with Spanish —which does not possess nasalized vowels— lost their nasalization at around the end of the Middle Ages, while the easternmost Souletin[6] —which is in close contact with the Gascon dialect of Occitan— is the only dialect that has maintained the opposition until today. However, it is worth noting that the other continental dialects (i.e. Labourdin and Low Navarrese) have lost the opposition, in spite of them all being in contact with French. It should also be considered that the aforementioned Biscayan and the peninsular dialects may have undergone different developments (and perhaps even chronologies) regarding this series.

4.3.3. Distribution of the Souletin /y/

Out of all the modern Basque dialects, only one, Souletin, has developed a vocalic system which diverges to such an extent from the reconstructed Proto-Basque vowels that it includes a sixth item. This item is the high front rounded

5 Text edited by Jeanne Viellard in 1938, *Le guide du pélerin de Saint-Jacques de Compostelle, texte latin du XIIe siècle* (Aimery Picaud, *Liber Sancti Jacobi*, *Codex Calixtinus*) (cf. *TAV*: 49-51).

6 Alongside its close but now extinct relative, the Roncalese.

vowel /y/. According to Michelena (*FHV*: 52), the realization of the S. /y/ is closer to the French /ø/ than it is to the French /y/. Hualde (1993b: 290) also describes it as being an intermediate sound between French /y/ and /ø/ and further notes, referring to Larrasquet (1932) and Lafon (1999 [1958]), that this fact is consistent with the fact that the other two high vowels of Souletin Basque (/i/ and /u/) are also more open than in French (Hualde 1993b: 290-291).

This vowel, which usually appears as <ü> in modern orthography,[7] developed from an older /u/ common to the other dialects and, although it can still easily be tracked back to that segment by means of cross-dialectal comparison, it is present in the very oldest documents written in this dialect.[8] The development of this vowel did not involve the loss of the older /u/ in Souletin, given that, although the /u/ fronting was almost context-free, it may have been hindered by certain consonants or clusters following the etymological /u/ (cf. infra). In addition to this, a later /u/ was also developed from /o/ in certain conditions, e.g. before a tautosyllabic nasal (cf. *FHV*: 54, and §4.3.4.2 here), after the fronting of the etymological /u/ was already complete.

The main difference between the /u/ fronting in Souletin Basque and the analogous processes in the neighboring Gallo-Romance languages is the context dependency which the Souletin process seems to exhibit. The contexts that block the change are not only far from being homogeneous or unifiable based on some phonological feature —this is not even the case when using a group of them— but they also seem to be very similar to other contexts in which the sound change systematically occurs. Michelena (*FHV*: 52-53) focuses on these seemingly similar sequences with different outcomes, previously compiled by Uhlenbeck (1909/10 [1903]), Gavel (1920) and Lafon (1999 [1937], 1999 [1958], 1999 [1962a], 1999 [1962b]). The following list is based on Oñederra's (2009b: 671) summary of the contexts mentioned by Michelena, who focused on those in which the sound change failed to take place (here reproduced under a.). Michelena also mentions some seemingly similar sequences of following consonants which did not hinder the process (under b.):

7 And as <u> in older ones, in which /u/ will usually be reproduced as <ou>.

8 It is present in Sauguis (Urquijo 1908/09) and Oihenart, who describes this vowel (Arnaud d'Oihenart, 1657, *Les Proverbes Basques recueillis par le Sr. d'Ohenart, plus les poésies basques du mesme auteur*, Paris), and represents it as <u>. It is also present in the first book written in Souletin (Jean de Tartas, 1666, *Onsa hilceco bidia*, Orthez), which is also consistent with this tradition (cf. *FHV*: 52).

a. contexts that do not yield fronting (*u* > *u*):[9]

a.1 / _ɾ (before tap): *gu(r)e* 'our', *gu(r)i* 'us (dat.)', *barur* 'fasting', *hirur (hiu)* 'three', *hur* 'water', *u(r)in* 'fat, grease'.

a.2 / _rð, rt, rtʰ (before rhotic-dental cluster): *urde* 'pig', *urdin* 'cloudy (water)', *urthe* 'year'.

a.3 / _s̻ (before apical fricative sibilant): *uste* 'supposition', *ikhusi* 'see', *busta*, *busti* 'wet'. But see also *üstel* 'rotten', *eüskara* 'Basque' and *deüs* 'something, nothing (when negated)'.

b. similar contexts that yield the change (*u* > *y*):

b.1 / _r (before trill): *hür* 'hazelnut', *ürrin* 'smell'.

b.2 / _rp(ʰ), rβ, rk(ʰ), rγ (before rhotic-labial or rhotic-velar clusters): *ürpho* 'manure pile', *bürkhi* 'birch', *ürgaiztü* 'help'.

/ _rl (before rhotic-lateral clusters): *khürlo* 'crane'.

/ _rs̻, rʂ, rʃ (before rhotic-sibilant fricative clusters): *ürsañ* 'sneeze', *ürzo* 'dove', *ürxantx* 'squirrel'.

/ _rh (before rhotic-aspiration clusters): *ürhe* 'gold', *ürhats* 'step'.

b.3 / _ʂ, t̻s̻, tʂ, (before laminal fricative sibilants and both apical and laminal affricates): *güzi* 'all', *hüts* 'void', *hütz* 'flatulence'.

As Oñederra (2009b: 666) observes, it is a very difficult task to find any feature that would unify those environments, and even more so to find any that would also give reasons for the blocking of the fronting process. Nevertheless, it seems rather plausible that the sound change has developed within the Basque language, given not only its unusual context, not present in any neighboring language, but also the fact that it does not show any affinity for loanwords (either from Gascon or from any other Gallo-Romance language) and thus cannot be straightforwardly explained as a borrowed development.

However, in the literature, the process has mainly been described as a result of the borrowing of this vowel from Occitan dialects, as proposed by Gavel (1920), Lafon (1999 [1962a], 1999 [1962b]) and Zarabozo (1972), among others. Hualde also speaks of it as undoubtedly resulting from its contact with Gascon (1993b: 290), but does not further specify as to its origin. Oñederra (2009a: 56-65, 2009b) takes a different stand and proposes an origin within the language itself, hypothesizing a rule telescoping. According to her, some gradual steps would have taken place between the /u/ and its actual outcome, in which a diphthongization and a subsequent monophthongization would have occurred

9 Lafon (1999 [1937]: 88) noted that the fronting also failed to apply before /k/ and /nk/ in the LN. dialect (variety of Mixe), and additionally before /g/ and /n/ in the variety from Bardoze (*FHV*: 53).

(2009b: 674).[10] However, she does not formulate any strong hypothesis —given the difficulty in linking the aforementioned contexts to the proposed development— nor does she explain the need for a diphthongization.[11] Against this hypothesis, crucially, stands the fact that there is no diphthong which is present in the other dialects but missing in Souletin, as may have been expected if they had yielded /y/.

The neighboring Gascon may in fact be relevant for the development of a split of the /u/ into /u/ and /y/, given that multilingualism or very close linguistic contact makes speakers more sensitive to subtle sub-phonemic contrasts, and that these contrasts may be prone to phonemicize over time (Blevins, p.c.). In this possible scenario, two different allophonic variants could have been present in the language, with back rounded vowels occurring before /ɾ/, /rð, rt, rtʰ/ and /s̻/, and their fronted allophones realized elsewhere.

When attempting to explain the contexts in which the different allophones are produced, certain details seem to be particularly pertinent, especially the division between rhotics.[12] Therein, the tap blocks the process, while the trill appears not to do so, with the exception of the trill/dental stop cluster, which is the only case in which the process is also blocked. Here, we could imagine a place assimilation which would make the coda rhotic a dental,[13] and we can wonder whether anything distinguishes this type of trill from the others. Indeed, Ladefoged and Maddieson point out[14] that trills usually have a lowered third formant, in contrast to uvular and dental *r*-sounds, which have a relatively high F3 (1996: 244). This could in fact be the phonetic cue favoring the non-fronting of /u/, given that the /u/ > /y/ change acoustically involves a lowering in the third formant along with the rise in F2.[15]

4.3.4. Vocalic processes

4.3.4.1. Alternations between /i/ and /u/, /e/ and /o/

As mentioned above, the only reason behind proposing a larger vocalic inventory for older stages of the language is based on the alternation between the high

10 Here, she follows Donegan (1978: 96).

11 If not by means of a phonetically more natural explanation, as understood within Natural Phonology.

12 Recall also that the block of the fronting preceding an apico-alveolar sibilant does not seem to be exhaustive.

13 As Michelena does (cf. *FHV*: 473).

14 Following Fant (1968).

15 This is a proposal which I intend to develop further in future works.

vowels. However, neither the quantity of the affected words nor their geographic distribution seem large enough to accept this hypothesis (*FHV*: 73).

There is a group of words with a general initial /i-/ which show an /u-/ in Biscayan and some other peninsular dialects instead of the /i-/: *igeri* : *uger* 'to swim', *ile* : *ule* 'hair', *(h)iri* : *uri* 'village', *iltze* : *ultze* 'nail'. Michelena (*FHV*: 280) suggests that the unprotected /u-/ may become /i-/ in disyllabic words when followed by an anterior vowel, and specifies that some intervening segments may block the assimilation, such as a trill or a rhotic-consonant cluster. There are some instances of this alternation which do not necessarily occur in word-initial position and which are also found in eastern dialects: *mutil* > R. S. *mit(h)il* 'boy'; *zubi* > R. *zibi* 'bridge'; *zumel* > HN. *zimel* 'withered'; *puxika* > LN. *bixika* 'bladder'.

The Roncalese dialect has regularly assimilated /i/ to /u/ when followed by /u/ in the following syllable, while Souletin has done so with /y/ instead, leaving the instances of /u/ unaffected (*FHV*: 79): *tipula* > **tupula* > *tupla* 'onion', *ilun* > *ulun*, *ülhün* 'dark', *irun* > *urun*, *ürün* 'spin (thread)', Sp. *hechura* > *itxura* > *ütxüra* 'appearance', *ekuzi* > *ikuzi* > *üküzi*[16] 'wash'. The reverse perseverative assimilation (/i/ to /u/ when preceded by /u/) also occurs in the same dialects, but in a sporadic manner: *burdina* > *burruña*, *bürdüña* 'iron', *urritz* > *urrutx*, *ürrütz* 'hazelnut tree'.

Alternations between /e/ and /o/ are not as usual as those involving high vowels, but Michelena also provides examples of some of these, especially in eastern dialects (*FHV*: 82-83): *sendo* > R. *sonto* 'strong', *behor* > S. *bohor* 'mare', *txekor* > HN. *txokor* 'young bull', *oren* > HN. *oron* 'for an hour'. According to Hualde, a similar assimilation is also found in the Biscayan variety of Bermeo (*GB*: 56): *etorri* > *otorri* 'come', *emon* > *omon* 'give', *erosi* > *orosi* 'buy', *eterno* > *otorno* 'funeral'.[17]

4.3.4.2. Different vowel risings

In addition to the diphthongization of mid vowels discussed in §4.3.5, there are some other, more common processes which result in a syllabic vowel becoming higher than before. These processes can affect any of the three non-high vowels in the Basque phonemic inventory and can occur not only by assimilation to a high vowel but also by the addition of a following vocalic segment.

16 The assimilation occurs after the mid vowel rising /e/ > /i/ (§4.3.4.2), as in the examples from Michelena (*FHV*: 79).

17 Hualde refers to Pérez-Bilbao (1991) in relation to this last example.

When preceded by a high vowel, /a/ has been raised to /e/ in multiple areas of the Basque Country: *ogia* > *ogie* 'the bread', *txakurra* > *txakurre* 'the dog', *katua* > *katue* 'the cat', *doguna* > *dogune* 'that we have', *izan* > *izen* 'to be', *ura* > *ure* 'water', *indar* > *inder* 'strength'. More precisely, the low vowel has been perseveratively assimilated to become a variant of /e/, which, according to Michelena (*FHV*: 47, 63, who mentions Navarro-Tomás 1923: 50), would be slightly more open than the standard. This process was already present in the 17th century.[18]

The high vowels have also anticipatorily assimilated a preceding mid vowel, especially in trisyllabic words. In such cases, /e/ and /o/ from the first syllable rise to /i/ and /u/: *hegi* > *higitu* 'move', *epistiko* > *ipiztiko* 'bishop', *mediku* > *midiku* 'doctor', Old Sp. *cobdiçia* > *gutizia* 'desire, whim' (cf. however, the vowel in **cupiditia* < Lat. *cupiditas*, *-ātis*), *ekusi* > *ikusi* 'see', etc. This process is more general and thus older than the progressive assimilation described above (apud *FHV*: 64).

A rising of the /o/ to /u/ occurs throughout the continental dialects, and is at its greatest strength in the Souletin dialect. There, this rising has developed regularly before a tautosyllabic /n/, being sporadic in the other dialects on the French side of the border: *gizon* > *gizun* 'man', *hon* > *hun* 'good', *untza* (cf. Sp. *onza*) 'ounce', *ezkondu* > *ezkuntü* 'married', *ontzi* > *untzi* 'container, ship'. This change is also present in some other barely specifiable contexts: *kozka* > *kuzka* 'hit', *molde* > *mulde* 'manner', *motz* > *mutz* 'short', *nor* > *nur* 'who', *nola* > *nula* 'how'.[19] Michelena disregards any connection to the similar development found in Bearnese, especially due to the fact that it is most present in the inherited lexicon and absent in certain borrowings which would be susceptible of developing the process, such as S. *botz* 'voice', in contrast to Bearn. *bouts* (*FHV*: 55).

Vowel rising develops regularly when any vowel is placed after a non-high vowel in inflection. This can occur both in nominal stems (with the addition of the article *-a*) and verbal stems (before the relative suffix *-en*, among others). The given contexts imply a late chronology for this process, since such vocalic clusters were only created after the loss of aspiration in the definite article during the Middle Ages.

The low vowel /a/ is raised to /e/ in Biscayan, Alavese and some Guipuscoan varieties under the aforementioned circumstances (*FHV*: 114-115):

18 As shown by the Biscayan text *Viva Jesus* (cf. Ulibarri 2010).

19 However, the vowel change in the interrogative pronouns could be a consequence of an analogy with the contextually conditioned *non* > *nun* 'where'.

neska-a > *neskea* 'the girl', *alaba-a* > *alabea* 'the daughter', *doa-an* > *doean* 'that goes', *daroa-anean* > *daroeanean* 'when (he/she/it) carries'.

The mid vowels also undergo rising when in stem-final position, after a following vowel is added to the stem in derivative processes. This rising is regular in many modern varieties of Basque and affects either /e/ or both /e/ and /o/ (apud Hualde *GB*: 47), as shown by the following examples: *etxe-a* > *etxia* (> *etxie*, cf. the aforementioned low vowel assimilation) 'the house', *gazte-a* > *gaztia* 'the young', *kafe-a* > *kafia* 'the coffee', *asto-a* > *astua* 'the donkey', *ard(a)o-a* > *ard(a)ua* 'the wine'.

Finally, there are some sporadic instances of /e/ > /i/ rising before /n/ or a sibilant which are mentioned by Michelena (*FHV*: 67-68), such as **e-nor* > *inor* 'nobody', *etxe* > *itxe* 'house', *etxola* > *itxola* 'shed', *(h)ertsi* > *itsi*, *itxi* 'close' and *pezu* > *pizu*, *pisu* 'weight'.

4.3.4.3. Vowel lowerings

A following trill, located either in the coda or at the onset of the following syllable, has opened /e/ to /a/.[20] This has occurred mainly in the western dialects, but only when /e/ was not at the very beginning of the word: *berri* > *barri* 'new', *txerri* > *txarri* 'pig', *gernu* > *garnu* 'urine', *piper* > *pipar* 'pepper', *maizter* > *maiztar*, *maistar* 'teacher' (< Lat. *magister*), *izter* > *iztar* 'thigh', *musker* > *muskar* 'lizard'. Nevertheless, there are also some examples of this lowering in the eastern dialects, such as *at(h)erbe* > *at(h)arbe* 'lodging', *pert(h)ika* > *part(h)ika* 'pole', *txert(h)atu* > *xart(h)atu* 'insert'.

Likewise, a tautosyllabic trill has sometimes turned /i/ into /e/ as well as /u/ into /o/ in the western dialects, or, at least, has produced some variation among these: *irten* > *erten* 'exit, leave', *kirten* > *kerten* 'handle', *bi(h)urtu* > *biortu* 'crooked', *agiri* > *ageri* 'manifest, evident'.

Finally, there are some variations between /a/ and /e/ or /e/ and /i/ before a tautosyllabic /l/, but there seems to be no such variation when they are located in the following syllable onset, as in common *beltz* > western *baltz* 'black', common *beldur* : western *bildur* 'fear' or eastern *elki* : *ilki* 'go/take out'.

4.3.4.4. Vowel addition

Prothesis is regular in loanwords beginning with a trill or a cluster with an initial sibilant. The prothetic vowel is usually an /e-/, but some instances of /a-/ are

20 Michelena notes that this occurs especially when /e/ is followed by a high vowel in the following syllable (*FHV*: 60).

found before /ra-/ or /ro-/. Any other vowel is the result of an assimilation (such as the /i-/ from /e-/ when preceding an /i/ in trisyllabic words, see §4.3.4.2). Examples of this process are: Lat. *rege* > *errege* 'king'; Lat. *ratione* > B. *errazoe*, G. *arrazoi* 'reason'; Sp. *raza* > *arraza* 'race', Sp. *ropa* > B. G. *erropa*, comm. *arropa* 'clothes', Lat. *spatha* > *ezpata* 'sword'; Lat. *spiritu(m)* (> **ezpiritu*) > *izpiritu* 'spirit, ghost'.

Other tautosyllabic clusters such as those composed of a stop and a liquid (present in Latin and Romance borrowings) have also been adapted to the phonotactic restrictions of the language by means of the addition of a vocalic segment. In this case, the epenthesized vowel is not a predetermined one, but its quality is predictable from the vowel found next to the "ill-formed" cluster: Sp. *Inglaterra* > *Ingalaterra* 'England'; Lat. *apricus* > *apirila* 'April'; Lat. *fronte* > *boronde* 'front'; Lat. *libru(m)* > *liburu* 'book'. Modern Basque is less restrictive insofar as these clusters are concerned.

In Guipuscoan, there are some instances of the addition of a palatal glide which is inserted between a vowel and a hetorosyllabic cluster formed by either a sibilant or a nasal and a following stop: *azken* > *aizken* 'last', *laster* > *laister* 'soon', *angeru* > *aingeru* 'angel', *aingira* (cf. Sp. *anguila*) 'eel'.

Much more common is the dialectal insertion of an epenthetic segment to break the hiatus between a high vowel and its following vowel. We find an instance of the development of an anaptyctic palatal glide between an /u/ and its following vowel in derived forms in the eastern dialects, as in *thüya* ['tʰyja] 'the spit' and *süya* ['s̻yja] 'the fire'. However, in Souletin and some varieties of Roncalese, the vowel sequence —[yja] and [wja], respectively— has later been simplified to *-ia* (cf. *FHV*: 121).

According to Bonaparte (1862: 33-34), the Biscayan variety of Orozko underwent an epenthetic process involving the insertion of /m/ between a stem-final /-o/ and a following vowel, which lasted until at least the 19th century. Examples of this include *ollo-a* > *olloma* 'the chicken' and *arto-a* > *artoma* 'the corn' (Hualde *GB*: 49).

A paragogic /-e/ is often developed in place names in the continental varieties as a consequence of a reanalysis of the vowel in locative suffixes, such as inessive *-(e)n*, allative *-(e)rat*, ablative *-(e)tik*, etc. as the end of the root. We find instances of this final *-e* in *Parise*, 'Paris' or *Miarritze* 'Biarritz', but also in words like *aurre* 'front' and *atze* 'back'.

In the Biscayan variety of Zeberio, a paragogic echo vowel is often added after a word-final *-k*, as reported by Etxebarria (1991). This is a copy-vowel, qualitatively similar to the preceding one, as shown in the examples: *zuk* > *zuku* 'you (erg.)', *nik* > *niki* 'I (erg.)', *gixonak* > *gixonaka* 'the man (erg.)'.

4.3.4.5. Vowel deletion

Apheresis is rare in Basque, although it can be found under certain specific circumstances. Firstly, there is a process of word-initial /e-/ deletion which occurs in different regions of Navarre, according to Hualde (*GB*: 57): *eman* > *man* 'give', *ekarri* > *karri* 'bring', *emazte* > *mazte* '(married) woman'. It can also result from morphological loss, principally of Latin prefixes, as stated by Michelena (*FHV*: 157, 511) and exemplified by Lat. *exemptus* > *sendo* 'robust'. Alternatively, it can result from a palatalization with a fusion of the /i/ with the new palatal segment, as in the modern *(t)xukatu* 'wiped away', whose stem Leiçarraga wrote as *ichuca*. However, this desegmentalization is mainly found in medial position and affects glides (*FHV*: 103), as shown by the much more usual examples: *zein* > *zeñ* 'who', *baino* > *baño* 'than', *(h)aize* > *axe* 'wind', *naiz* > *nax* '(I) am'.

Vowel deletion is particularly common in hiatus (see §4.3.5). The combination of /a/ followed by any other vocalic element often leads to the loss of the /a/: *da+en* > *den* '(that it) is'; *za+en* > *zen* '(that it) was'; Old B. *Bilbao* > modern *Bilbo* 'Bilbo'; *hamabortz* > L. *amortz* 'fifteen'; **baga-uso* > Al. *baguso* 'ring dove', **aldairi* > *aldiri* 'proximity'. For the sequence of a high vowel followed by any other vowel in medial position, the latter (if any) will be deleted: *amuarrain* > B. *amurrai(n)* 'trout', *luebaki* > *lubaki* 'trinchera', *bariaku* > *bariku* 'Friday', *azienda* > *azinda* 'livestock'. When two mid vowels are combined, /o/ remains while /e/ is usually dropped, as in *moeta* > *mota* 'kind' and *adiskideok* > B. *adiskidok* '(these) friends'. The deletion of the second vowel in a /VjV/ sequence is very common: *baieta* > *baita* 'too', **bai-ez-ik* < *baizik* 'but rather', *jaiotegun* > *jaitegun* 'day of birth'.

In some dialects, an intense process of syncope is found. It is particularly widespread in the eastern regions of Salazar and Roncal —but not in Souletin—, although it also occurs in High Navarrese, Low Navarrese and Aescoan (*FHV*: 160). This syncope develops in the following contexts, as described by Michelena (ibid.):

Contexts that yield syncope:

a. / T_ɾ, l (between a stop and a tap or a lateral, mainly in R and Sal but present in other dialects): comm. *aingeru* > *aingru* 'angel'; comm. *denbora* > R. *tenpra*, Sal. *tenbra* 'time'; S. *nabela* > Sal. *nabla*, R. *ñabla* 'knife'; LN. *gira* > R. Sal. *gra* 'we are', comm. *andere* > HN. LN. R. Sal. *andre* 'lady'
b. / S _ɾ (between a sibilant and a tap, only in R and Sal): LN. *zira* > R. Sal. *zra* 'you are', LN. *ziren* > R. Sal. *zren* 'they were'.
c. / ɾ, r_O (between a rhotic and an obstruent): L. LN. *eriden* > R. Sal. *erden* 'find'; comm. *eraman* > Sal. *erman* 'carry'; comm. *baratze* > R. *bartze*

'vegetable garden'; comm. *-arena* > R. Sal. *-arna* 'the one from'; comm. *-garrena* > R. Sal. *-garna* 'the Xth'.

d. / O_O (between two obstruents): Aesc. *zitu* > R. Sal. *ztu* 'has (...) you' (cf. also the plural R. *ztei*, Sal. *ztie*); Aesc. *tuzu*, *tuzie* > Sal. *tzu*, *tzie* 'you have (sg./pl.)'; comm. *ganibeta* (> **gabineta*) > Sal. *gaminta* 'knife'.

These contexts include virtually any outcome which the phonotactics of the dialect could handle: e.g., no Basque dialect favors nasal/tap clusters and thus these are absent, while tautosyllabic clusters involving a sibilant only appear in the varieties from Roncal and Salazar.

There is a final syncope process which results in the loss of an /e/ in the second syllable and after a trill. This is a general development which is not so dialectally biased as the one described above. Examples of this syncope are mainly found in loanwords that have already developed a prothetic /e-/, such as *erreloju* > *erloju* 'clock', *errelijio* > *erlijio* 'religion', *errelazio* > *erlazio* 'relation', Sp. *reyerta* > *errie(r)ta* 'brawl', etc.

Apocope is a common device in composition, in which it is systematically applied to the first member of any trisyllabic word ending in a vowel. Examples of this process are Ax. *burdin gori* 'red-hot iron' (cf. *burdina* 'iron') and *(h)euskaldun* 'Basque speaker' (cf. *euskara* 'Basque'). For disyllabic words, /-i/ is regularly lost while /-u/ is only dropped in eastern dialects, and non-high final vowels are systematically neutralized to /-a/: *hardi* 'stony area', *zaldun* 'knight', *burhezur* 'skull', *urdaki* 'pork fat' or *gogatsu* 'enthusiastic' from, respectively, *harri* 'stone', *zaldi* 'horse', *buru* 'head', *urde* 'pig' and *gogo* 'will, mind' (*FHV*: 125-127).

Besides the ones involved in composition, there are some words which can undergo apocope. In Guipuscoa, some words lose their final /-a/ (which corresponded to the root) after another vowel, as a consequence of its reinterpretation as the definite article: *burnia* > *burni* 'iron', *katea* > *kate* (cf. Lat. *catēna*) 'chain', *koro(i)a* > *koro(i)* (cf. Lat. *corōna*) 'crown'. Some other words such as *barda* > *bart* 'last night' or *giltza* > *giltz* 'key' also occur in different varieties, as well as the widespread verbal forms such as **naiza* > *naiz* '(I) am'.

Finally, we find a (very modern) process of apocope in the Biscayan varieties surrounding Markina and Getxo. In this process, the second vowel (the *-a* or *-e* from the definite article) of an inflected nominal stem is regularly dropped by young speakers, even in final open syllables.[21] Examples of this include: *neskie*,

21 Hualde (*GB*: 50) reports that this second vowel loss is much more geographically widespread in closed syllables than it is in open syllables.

neskea[22] > *neski*, *neske* ‘the girl’; *alabie*, *alabea* > *alabi*, *alabe* ‘the daughter’; *kafie*, *kafea* > *kafi*, *kafe* ‘the coffee’; *astue*, *astoa* > *astu*, *asto* ‘the donkey’.

4.3.4.6. Vocalic reciprocal metathesis

Vocalic metatheses are not rare in the history of the language, but the process is conditioned by a common phonological feature shared by both of the reordered segments, so that only certain possibilities of transposition are attested. The relevant feature which constrains two-vowel non-local metatheses is [+/- high], which divides the 5-vowel inventory into two groups of segments. A given segment can change position with another one only if both are members of the same group, i.e. reciprocal metatheses only occur among segments that agree in height (Egurtzegi 2011). Examples of metathesis of non-high segments include: *atera* : *etara* ‘come out, take out’, *alkandora* : *alkondara* ‘shirt’, and *(h)odei* : *(h)edoi* ‘cloud’, while *ikutu* : *ukitu* ‘touch’ and *ingude* : *ungide* (< Lat. *incūde*) ‘anvil’ exemplify metatheses involving a pair of high vowels.

4.3.5. Diphthongs

The diphthongs which Michelena proposes for Proto-Basque are /au̯/, /eu̯/, /ai̯/, /ei̯/ and /oi̯/ (*FHV*: 87). Regardless of their common status in modern Basque, diphthongs seem to be secondary (probably in their entirety), and a significant rise in their frequency can be traced back. Diphthongs usually originate from the loss of an approximant, aspiration[23] or tap,[24] etymologically located between the two vowels. Given that most of these processes probably occurred some centuries after the Michelenian Proto-Basque, it is very likely that only some of the aforementioned five diphthongs were known at this stage, if any.

Diphthongs can also be nasal in the varieties in which the intervening source of the nasality is lost and a vocalic nasality contrast is still productive, as in Roncalese or in 16th and 17th century Biscayan. Although falling diphthongs (/ai̯/, /au̯/, etc.) are more widespread than their rising counterparts (/ja/, /wa/, etc.), some of the latter may be older than the former, at least those involving a word-initial yod.

One of the reasons behind the older status of rising diphthongs is that their word-initial position, in an Old Proto-Basque mainly consisting of roots and

22 The variety from Getxo does not rise the prevocalic mid vowels as in Markina (cf. §4.3.4.2).

23 Aspiration which was a former nasal, cf. §4.4.3.1.2.

24 In more modern stages of the language than the others.

their prefixes, was peripheral to the root as well as to the stress. Thus, word initial segments were the most susceptible to suffering different kinds of deletion, including those deletions yielding consecutive vowels that would later become diphthongs. These kinds of yod-based vocalic clusters are usually found in verbal forms and they originated from the combination of an initial /e/ and a following vowel, which were most probably separated by a coronal approximant in older stages of the language. Examples of this process include: **e-da-kin > jakin* 'know' and **eoan > joan* 'go' (according to Michelena *FHV*: 119).

4.3.5.1. Falling diphthongs

Falling diphthongs are formed by the encounter of a non-high and a following high vowel, either in composition or derivation, and after the occurrence of a phonological process affecting an intervening consonant. /a/ may combine with either of the glides, although these are frequently reduced to a simple high vowel in composition (see *FHV*: 116), as in *basa-urde > basurde* 'wild boar' and *garaunak > garunak* 'brains'. Examples of this combination include those created after the loss of the aspiration, as in *ahuntz > auntz* [au̯nt̪s̪] 'goat', and *nahi > nai* [nai̯] 'will', which was still disyllabic in RS.

The mid vowels are normally combined so that diphthongs are formed by anterior/posterior or posterior/anterior pairs (cf. ibid.), as in *ehun > eun* [eu̯n] 'hundred' and *ohi > oi* [oi̯] 'habit'. The combination of /a/ and a mid vowel may also result in diphthongization (*apaez > apaiz* 'priest', *hama-eka > hamaika* 'eleven'), but the most usual outcome would involve a contraction favoring the latter element.

The general tendency noted by Michelena (*FHV*: 486) implies that word-initial falling diphthongs usually lose their second element. He also emphasizes that any diphthong ending in a palatal will lose its second element before a rhotic, as in *sorburu* 'back' from *soin* 'shoulder' and Lç. *ohorgoa* 'larceny, theft' from *ohoin* 'thief' (cf. *jaurgoa* 'empire' from *jaun* 'Sir').

Along the same lines, a palatal off-glide is often lost after the palatalization of a following consonant. This is the consequence of a reinterpretation of the source of this feature as originating in the consonant instead of the etymologically palatal vowel. Examples of this are mainly found in the Biscayan dialect —as well as in peninsular areas— but can be found in any palatalizing variety: *naiz > nax* '(I) am', *noiz > nox* 'when', *baina > baña* 'but', etc.

There are some geographically widespread cases of the dissimilation of an off-glide, when followed by its equivalent vowel in the following syllable. /w/ can dissimilate to /j/, as in LN. S. *haizu* < Lat. *ausus (sum)* 'to be lawful, to dare'; *errekaitu* 'provision, food' < Sp. *recaudo*, but they are usually dropped:

auntzume > B. G. *antzume* ‘kid (goat)’; *itaundu* > B. *itandu* ‘to ask’; *jaiki*, *jeiki* > *jaki*, *jeki* ‘stand up’; *-gaitik* > *-gatik* ‘because of, due to’.

Syncope has spontaneously affected /au̯/, simplifying the diphthong after the loss of the off-glide. The frequency of this process increases with the length of the word, and is particularly usual in initial position and among verbal forms: *aulki* > *alki* ‘chair’, *laurogei* > *larogei* ‘eighty’, as well as in Lat. *audītum* > *aditu* ‘listened, understood’, provided that the etymology is correct. Other simplifications are rare, at least in initial position (cf. *FHV*: 491).

/ai̯/ is relatively stable, although the low vowel can sometimes be assimilated to /e/, even resulting in its loss (cf. infra for /ei̯/ > /i/): *gai* > *gei* ‘matter’, *jaiki* > *jeiki* ‘wake up’, *egitai* > *igitei* > *igiti* ‘sickle’. In Souletin and Roncalese, an older **au* has regularly become /ai̯/, except when followed by /ɾ/, /r/, /s̻/ or /ts̻/: *gau* > *gai* ‘night’, *gauza* > *gaiza* ‘thing’, *laudatu* > *laidatü* ‘to laud’.

/eu̯/ is much more unstable than /au̯/, and different processes have affected it so that nowadays it is rare. It has been simplified by the loss of either of its elements —usually /e/, as in *heuragi* > *ugari* ‘abundant’ and G. B. *euli* > HN. LN. L. *uli* ‘fly’ (but cf. also the diminutive *eltxo* ‘mosquito’)— or dismantled by means of excrescence, as in *leun* > *legun* ‘soft’ and *neurri* > *negurri* ‘measure’.

The stability of /ei̯/ is largely dependent on its position within the word. In initial position, apheresis has regularly occurred, with the initial /ei̯-/ being practically absent from modern Basque: **eizan* > *izan* ‘to be’, *eihar* > *ihar* ‘dry’, although in the latter the semivowel may also be epenthetic (cf. Lakarra 2009c).[25]

/oi̯/ is usually maintained, although it sometimes alternates with the rarer /wi/ (< /ui̯/), as in the collective suffix *-doi*, *-dui*, *-di* from toponymy. This diphthong sometimes comes from /ei̯/, as in *hogei* > LN. L. *hogoi* ‘twenty’ or Aesc. *oreitu* > *oroitu* ‘remember’.

4.3.5.2. Rising diphthongs

Rising diphthongs develop in almost the same circumstances as their falling counterparts, but these are particularly habitual in the nominal forms of certain verbs. In these cases, as mentioned above, an initial **eðV-* becomes *jV-* after the loss of the intervening consonant. Such forms are usually followed by /a/ due to the high frequency of an older **e-da-*,[26] as in *jakin* ‘know’, *jausi* ‘to fall’, *jaun* ‘sir’, etc. It is also frequent before /o/ (*jo* ‘hit’, *josi* ‘sew’, *joan* ‘go’) and much rarer before /i/ (*jin* ‘come’). The examples of /je-/ are much more modern, and

25 On the other hand, in medial and final position, /ei̯/ is more stable, although it can also be dropped there.

26 On this second morpheme, see Lakarra (2006a, 2006c).

they continue —perhaps in their entirety— from an older **ea-* > **ja-*, given that an etymological **e-e-* may have yielded *e-*. Examples of this include: *jasan* > *jesan* 'borrow', *jaitsi* > *jeitsi* 'go down' or *jasarri* > *jesarri* 'sit down'. This initial yod is absent before /u/, maybe as a result of the formation of the falling diphthong from **eu-* sequences, instead of the raising diphthongs seen in the former vowel combinations (i.e., **eu-* > *eu-* and not > ***ju-*). We can reconstruct an example of this: **e-dun-tz-i* > *eutsi* 'hold'.[27]

Rising diphthongs involving /w/ have usually been simplified, yielding different results: /wa/ has been replaced by /o/ (especially after a velar stop), as in *kodaña* (cf. Sp. *guadaña*) 'scythe' and *gorde* (cf. Sp. *guardar*) 'save', and /we/ has been syncopated to /e/, as in Sp. *ruego* > *erregu* 'request' and perhaps in *lueco* > *leku* 'place' (*FHV*: 167-168).

4.4. Consonantism

4.4.1. Sonorants

4.4.1.1. Rhotics

Most modern Basque dialects oppose an apical tap to a trill, also apical. Nevertheless, the opposition among rhotics is only appreciable between vowels, given that it is absent from initial position and has been neutralized to the trill in final (in every dialect except Roncalese) and in pre-consonantal position.

In the native lexicon, rhotics do not occur in initial position, and a prothetic vowel is systematically added in loanword adaptations. This vowel may be either /e/ or /a/ (see §4.3.4.4 on vowel addition).

A straightforward consequence of this may be the removal of the onset rhotic from the oldest stages of the language that we are currently able to reconstruct, i.e. we can hypothesize that a monosyllabic Basque root had only word-final —maybe non-contrasting— rhotics. Later, alveolar taps would result from the intervocalic /l/ in a regular manner (cf. §4.4.1.2), while coda rhotics would slowly evolve towards the trill, creating an intervocalic opposition after the loss of the consonant which followed the coda trill, and its subsequent resyllabification. Pre-consonantal rhotics have systematically yielded a trill until very recent times, and it is possible that this was the case in Proto-Basque. Modern examples of these are found after a resyllabification caused by the loss of a following onset, as in the dissimilations or metatheses in the Biscayan names *Fernando* > *Ferrando* and *Gernika* > *Gerrinka* or *Zornotza* > *Zorrontza* (cf. Egurtzegi 2011),

27 Cf. Lakarra (2006a) on the -(*n*)*tzi* > -*tsi* development proposed by himself.

and the loss of the aspiration on the Labourdin coast: *nerhabe* > *nerrabe* 'single man', *orhoitu* > *orroitu* 'remember' and *erhaztun* > *erraztun* 'ring' (cf. the unaspirated common forms *nerabe*, *oroitu* and *eraztun*).

When they represent the second element of an onset cluster, rhotics seem to be historically perceived as a tap. Given that Proto-Basque disallowed such tautosyllabic clusters, we find evidence for this assumption in the adaptation of Latin borrowings such as *libru(m)* > *liburu* 'book', *escribe(re)* > *izkiri(b)atu* 'write' or *fronte(m)* > *boronde* 'front' or *lucru(m)* > *lukuru* 'profit, gain'.

As stated above, the distinction between rhotics is not made in the coda, as most Basque varieties only produce a trill in this position. However, as may be inferred from historical evidence, this opposition was in fact effective before an aspiration,[28] although this can no longer be discerned since the aspirating dialects have lost their rhotic opposition due to independent reasons. However, it is most probable that this contrast was lost even before the rhotic neutralization.

As a consequence of its intervocalic tap loss, modern Souletin has only a single rhotic (*FHV*: 328-330). However, this is only the case for recent times, since the tap has been lost in the course of the last two centuries and has played a role in many phonological changes. The remaining rhotic is produced as an alveolar trill, although it has fewer vibrations than the common Basque trill (Hualde *GB*: 30-31).

Also in very recent times, in most of the varieties from Labourd and Low Navarre, the apical articulations have merged into a posterior realization. Nevertheless, Gavel reported that, in sporadic cases, some speakers have conserved an apical tap along with the uvular trill (1920: 239). The trill first became uvular, and then the tap began to be produced likewise, giving rise to their merger in most cases. The rest of the dialects[29] maintain both rhotic segments intervocalically, although /ɾ/ tends to drop sporadically in many varieties (cf. *FHV*: 330).

The process of neutralization in final position seems to be complete nowadays, after the disappearance of Roncalese, the only dialect which discerned between the two apical rhotics word-finally (cf. Azkue 1931: 218-219). However, there is still a small group of words which possess an underlying word-final tap that surfaces in derivation: *nor* 'who', *zer* 'what' (alongside *inor* 'nobody' and *ezer* 'nothing'), *(h)ur* 'water', *zur* 'wood' and *hor* 'dog'.[30] The non-fronted high rounded vowels in Souletin *hau* 'this', *hiru* 'three' and *lau* 'four' denote that the

28 Hualde only mentions the example of Souletin as possessing a contrast before the aspiration, but Leiçarraga also shows a systematic differentiation between rhotics in this position. He writes *urrhets* 'step', *bur-hezur* 'skull', etc.

29 Those located in Spanish territory.

30 A couple of recent loanwords increase the number of words in this group: *paper* 'paper' and *plater* 'dish'.

final **-r* reconstructed for those forms (cf. the older *haur*, *hirur*, *laur*) was in fact a tap.[31] Some of these words, especially *(h)ura* 'the water', also have dialectal variants involving palatal realizations which are very different to both the etymological tap and the otherwise extended word-final trill. Such examples include the voiced affricate and approximant in [uʤa] and [uja] from the Guipuscoan varieties of Etxarri-Aranatz and Goierri, respectively (cf. *FHV*: 556).

There are some instances of compounds in which an etymological final tap is found as an aspiration in the first member of the compounded form. This process seems to develop in old monosyllabic words, such as *hur* 'water', *zur* 'wood' and *hor* 'dog': *uhalde* 'shore, riverside', *uharte* 'island', *ühatx* 'torrent', *zuhaitz*, *zühaintze*, *zühañ* 'tree', *zuhandor* 'shrub, bush', *zübihotz* 'center of the oak', *ohara* 'dog in heat', *ohalano* 'mastiff', etc. This process seems to be common, and thus prior to dialectal developments such as the vowel fronting found in the Souletin variants above, which would not have occurred before a tap.

4.4.1.2. Laterals

Common to all modern Basque dialects are the apico-alveolar and palatal[32] laterals, but another one —usually transcribed as **L*, following the Michelenian tradition— is postulated for older stages of the language. Although the most widespread realization of /l/ is apical, Michelena (*FHV*: 311) reports the presence of velar productions in coda position in some Souletin territories, as well as isolated retroflex productions (*FHV*: 550).

During the Middle Ages —before the 11th century, since this evolution was already completed in the first written documents— the alveolar lateral yielded /ɾ/ intervocalically, much in the same way that /n/ had previously changed to /h/ (cf. §4.4.1.3.1 and §4.4.3.1.2). Although some exceptions can be found, this process occurred in a regular way: Lat. *angelu(s)* > *aingeru* 'angel', Lat. *gula* > *gura* 'hunger, desire', Lat. *voluntate(m)* > *borondate* 'will'. Most exceptions either correspond to recent borrowings or show dialectal variation (e.g. B. *solo* : comm. *soro* 'field, meadow' (cf. Lat. *solum*, Sp. *suelo*) and comm. *-(a)ra* : S. Sal. *-ala* 'to', although the latter could be derived from two suffixes: **-an-la* > **-alla* > *-ala*).

This process created an alternation between a name and its composed or derived forms, after the loss of the final vowel[33] had bled the phonological envi-

31 See §4.3.3 on the Souletin /y/.

32 On the development and hypocoristic usage of /ʎ/, see §4.4.4.1.

33 Which is systematic for trisyllabic words in composition, cf. §4.3.4.5.

ronment when followed by an initial consonant in the second member of the compound: *euskara* 'Basque' > *euskaldun* 'Basque speaker', *erdara* 'foreign language' > *erdaldun* 'foreign speaker', etc.

The fortis **L*, a segment reconstructed by Michelena for Proto-Basque, parallels the rest of the sonorant fortes *R* and *N* (cf. §4.4.1.1 and §4.4.1.3) respectively. The primary reasons for the reconstruction of this segment stem from the high frequency of the modern intervocalic /l/. Recall that **l* systematically evolved to /r/ in this position, as well as the observation that Latin *-Vl.lV-* had regularly yielded Basque *-V.lV-* (cf. *ballaena* > *balea* 'whale', *cella* > *gela* 'room' and *castellum* > *gaztelu* 'castle'). Thus, Michelena proposes a lateral segment that would ultimately become the modern intervocalic /l/ for all the reconstructions which share the outcome of this development, present in Latin loanwords.

Although it remains uncertain as to whether this segment had a different realization[34] or was solely a longer production of the alveolar /l/ (i.e. a geminate), what seems to be clear is that any /l/ which was not found in an intervocalic position at the time of the **l* > *r* change would remain the same after a late alteration of its etymological phonological context. Consequently, any lateral historically located before an aspiration appears as an /l/ in all modern dialects[35] (cf. *alhaba* > *alaba* 'daughter') as does any lateral which etymologically belonged to a cluster which may be absent today (cf. **erlana* > **erlaha* > *elae* > *elai* 'swallow (bird)').[36]

This segment appears in the *Reja de San Millán* (*CSM*, 1025) as both *-lh-* and *-ll-*, and, according to Michelena (*FHV*: 321), this seems to be related to the restrictions applied to the aspiration within the boundaries of the word, i.e. /Vl.hV/ occurs between the initial and second syllables and /Vl.lV/ elsewhere. It is also emphasized by Michelena that in order to justify the high number of modern word-final laterals, the lateral segments may have been limited to **L* word-finally, since they never become taps before the definite article (e.g. *Kessalla* > *gezal* 'salted water' and *Zaballa* > *zabal* 'broad', cf. *FHV*: 322).

However, there may be another explanation for this phenomenon, since the definite article, which is reconstructed as **-(h)a(r)* and present as *-ha* in San Millán (cf. Manterola 2006: 674), together with the distribution of the /lh/ and /ll/ within the frame of the word mentioned above, lead us to think that heterosyllabic *-l.h-* clusters may have yielded *-l.l-* (which would later simplify to *-.l-*)

34 Michelena (*FHV*: 320-321) suggests that **L* was a segment —and not a cluster— in the first centuries of the Middle Ages.

35 *-lh-* in the aspirating dialects and *-l-* in the dialects that no longer possess it.

36 See Ariztimuño & Egurtzegi (2011) on this etymology.

after the loss of aspiration of the definite article. This explanation does not necessarily rely on the dropping of the aspiration after the second syllable (cf. §4.6.4), since the process of grammaticalization itself would explain the loss of the /h/. This explanation, in contrast to that of final or boundary lengthening, would not hinder reconstructions involving the *-*l* > *ɾ* change in suffixed verbal forms which add an *-i* to a root ending in a lateral, such as **e-dol-i* > *jori* 'abundant, plenty' or **e-lol-i* / **e-ra-dol-i* > *erori* 'fall' (cf. also the related reduplicated form *odol* 'blood', apud Lakarra 2011b).

There are some word-initial laterals that came from an older /d-/ in words such as Lat. *thēca* > Basq. *leka* 'pod', **doni* > modern *lohi* 'body', etc. This process may have been systematic in non-verbal words, given that most initial /d/s are found in auxiliary and synthetic verb forms, along with recent loanwords in modern Basque (cf. §4.4.2.1).

4.4.1.3. Nasals

In modern Basque, an alveolar /n/, a bilabial /m/ and a palatal nasal /ɲ/ contrast prevocalically, but only the alveolar and the palatal segments contrast word-finally.

When in word-medial coda, nasals are systematically assimilated to the following consonant and are realized with the same place of articulation. Examples of this assimilation are: *enbor* [embor] 'trunk, log',[37] *handi* [haⁿ̪di] 'big', *onddo* [oɲɟo] 'fungus', *aingura* [aiŋgura] 'anchor', *anfibio* [aɱfibjo] 'amphibian', etc.

Along with the nasal stops, a nasalized glottal fricative /h̃/ is also present in modern Souletin (see §4.4.3.1.2).

4.4.1.3.1. /n/

Although the modern language possesses three contrasting nasal obstruents in all of its varieties, historical and, to a greater extent, reconstructive evidence suggests that only the alveolar obstruent was present in the oldest stages we can reach. However, not every alveolar nasal has remained unchanged until today.

As a matter of fact, Proto-Basque /n/ was regularly affected by a rhinoglottophilia process (cf. Matisoff 1975, Igartua 2008) which altered it to the nasalized aspiration /h̃/ in intervocalic position, probably its most usual location (cf. *FHV*: 142-153, 215, 300-303; Igartua 2011 and §4.4.3.1.2). Since an older etymological /h/ was also present between vowels, both aspirations contrasted in this position, and perhaps even prevocalically to a lesser extent (cf. §4.4.5.3 on

37 Which does not yield ***emor*, cf. §4.4.1.3.2.

consonantal metathesis). However, /h̃/ only remains in the Souletin dialect nowadays, since it has lost its nasalization —and thus merged with /h/— in the rest of the aspirating varieties.

After the complete loss of the aspiration in the western dialects, the nasalization was kept in the vowels etymologically surrounding the nasal segment for some centuries, so that it still appears in some Biscayan documents from the 16th and 17th centuries (cf. §4.3.2 on nasalized vowels). This nasalization was often reanalyzed as a syllable-final obstruent, which was subsequently restored after the diphthong: **artzani* > **artzaĥi* > *artzãĩ* > *artzain* 'shepherd', Lat. *lūcānica* > **lukaĥika* > **lukãĩka* > *lukainka* 'spicy pork sausage', **arrani* > **arraĥi* > *arrãĩ* > *arrain* 'fish', Aquit. *seni-* > *seĥi* > Old B. *sẽĩ* > modern *sein* 'child'.

The reestablishment of the nasal stop is very common after an /i/, but a palatal nasal is reinstated instead of the etymological apico-alveolar: comm. *katea* : S. *kathiña* 'chain', Al. *erregia* : B. G. *erregiña* 'queen', L. *ihes* : B. *iñes*, etc.

An intervocalic /n/ usually yields /ɲ/ when following an /i/ or a yod, as described in more depth in §4.4.4.2. In many cases, the palatal off-glide merges with the following nasal, so that the etymological context which produced the change is regularly lost in some varieties and thus creates a contrast between the two nasal segments intervocalically.

Along the same lines, there are also some isolated examples of the restitution of an intervocalic /m/ after an /u/: L. LN. *zuhai* : HN. *zumai* 'forage'; L. LN. *zuhar* : HN. G. B. *zumar* 'elm'. An /n/ can also yield /m/ after an /u/, as in *un(h)il* > L. *umil* 'funnel'.

4.4.1.3.2. The development of /m/

Despite its modern high frequency, the Proto-Basque consonantal system lacked a labial nasal, as was made apparent by Michelena (*FHV*: 270) and accepted by Trask (*HB*), who also pointed out its lack of morphological value. The segment was later introduced by means of borrowing from Latin and Romance languages and also developed within the language itself.

Two different assimilatory processes which led to the nasalization of a labial (oral) stop are reconstructed with different chronologies. Common to all of the dialects is the transformation of a nasal and a labial stop sequence into a labial nasal: /nb/ [mb] > /mm/ (> *m*). This development would have occurred in the first centuries of our era, since *Sembe-* (cf. modern *seme* 'son'), *Ombeco* and

Ombexonis (usually linked to *ume* 'child'), attested in Aquitanian, do not show the assimilation of the labial stop.[38]

The second process of assimilation, the similarly widespread regressive assimilation of nasality, has been productive within the modern dialects themselves. In this latter process, a non-local assimilatory nasalization of /b/ occurs when followed by a nasal obstruent (cf. §4.4.5.1), as in *bendekatu* > *mendekatu* (cf. Lat. *uindicāre*) 'avenge', *bagiña* > *magi(ñ)a* (cf. Lat. *uāgīna*) 'scabbard, sheath' or *Mendekoste* (< Sp. *Pentecostés*) 'Pentecost'. Although we can extend its productivity to some centuries onwards, the latter assimilation also occurred before the rhinoglottophilia process had apparently ended. This is shown by the chronology necessary in multiple standard reconstructions such as **bene* > **mene* > *mehe* 'thin' or **bini* > **mini* > *mihi* 'tongue', although the assimilation of nasality may also have originated from a nasalized aspiration.

There are nevertheless some phonological contexts which block the assimilation of nasality (cf. Egurtzegi 2011, §4.4.5.1), which would probably involve the presence of an intervening segment such as a coda /r/ between the /b/ and the following /n/. This is exemplified by the form *tramena* 'tavern', which shows an assimilation, in contrast with other variants such as *taberna*, *taferna*, etc., from which it is consistently absent. The apparent lack of ***mVrn* sequences backs this hypothesis. The biggest problem for identifying the aforementioned contexts lies in the large number of examples in which a voiced labial stop has become nasal even when it was not followed by a nasalized environment. Examples of this /b/ - /m/ alternation are: *makulu* (< Lat. *bacculus*) 'walking stick', *makar* < *bekar* 'sleep, rheum' and *maru* (< **balu* < Lat. *pālus*) 'pole, post'.

As a consequence of its marginal status in older stages of the language, /m/ has had an expressive value similar to that of the palatal segments (cf. §4.4.4.1), and it appears in the second component of reduplicated forms such as *zaldiko-maldiko* 'carousel', *zirimiri* 'drizzle' and *xistmist* (> *tximist*) '(lightning) bolt'.

4.4.2. Stops

All modern Basque dialects have at least two series of stops (voiced and voiceless, alongside aspirated voiceless stops in the eastern varieties) which are produced at four different points of articulation: labial, dental, palatal and velar. Palatal stops are less frequent than the others and often serve a very specific affective purpose (cf. §4.4.4.1).

38 Unlike other Aquitanian forms such as *Ummesahar* (cf. also modern *ume* 'child') from Lerga, but this form also shows the expected vocalism, in contrast to *Ombeco* and *Ombexonis* mentioned above.

Within the modern dialects, stops are mainly found in onset position, since they are limited to the borrowed vocabulary in preconsonantal position, and only /-t/ and /-k/ are (very infrequently) found word-finally. These word-final stops are the result of apocope followed by a final devoicing (cf. §4.4.2.2), and there is no reason to think that stops could have appeared in coda position in older times.

4.4.2.1. Voiced stop series

Basque voiced stops can be produced either as stops or as approximants. Stops are produced phrase-initially and in certain postconsonantal positions. This is systematic for any voiced stop located after a pause as well as after a nasal and, in the case of /d/, it is also realized as a stop after /l/. In those contexts in which the spirantization of the voiced stop does not occur, the place of articulation of the preceding segment is assimilated to that of the stop (cf. §4.4.1.3): *enbor* [embor] 'trunk, log', *handi* [haɳdi] 'big', *iñddar* [iɲɟar] 'strength', *aingura* [aiŋgura] 'anchor', *alde* [alde] 'side'. Approximants are found between two vowels, as well as between a heterosyllabic continuant consonant (/l/ —with the previously mentioned exception of [l̪.d]—, /r/, /s̪/ and /ʂ/) and a following vowel): *arreba* [areβa] 'sister (of a boy)', *adostu* [aðos̪tu] 'agree', *lagun* [laɣun] 'friend', *albo* [alβa] 'side', *elgar* [elɣar] 'each other', *arbaso* [arβas̪o] 'ancestor', *ardo* [arðo] 'wine', *argi* [arɣi] 'light', *ezberdin* [ez̪βerdiɲ] 'different', *ezdeus* [ez̪ðeus̪] 'nothing', *desgogo* [dez̪ɣoɣo] 'reluctance', etc. Note that, as in Spanish, fricatives are voiced before a voiced consonant in modern compounds (cf. §4.4.3.3.2). However, stops have been regularly devoiced after sibilants in the native vocabulary as well as in old borrowings and compounds, so that contextually voiced sibilants are rare before stops.

This process of spirantization has sometimes even yielded the complete loss of the etymological voiced stop, although this has only occurred in intervocalic position and not after a continuant. Given that intervocalic approximant loss is found in onomastics dating from at least the 12th century, the spirantization of intervocalic voiced stops must have been even older. Old examples of this loss are usual in compound surnames such as: *Soracoiz* (< *sora-bakoitz*), *Sorauren* (< *sora-guren*), *Echarri*, *Echerri* (< *etxa-barri* / *-berri*), etc.

Instead of yielding deletion, some instances of intervocalic /d/ (more accurately, [ð]) have become taps in many Biscayan, Guipuscoan and High Navarrese areas. This process may appear to be almost systematic in some dialects, but [ð] has not merged with /ɾ/ anywhere (apud *FHV*: 228): *edo* > *ero* 'or', *edan* > *eran* 'drink', *odoi* > *oroi* 'cloud', *etorri re* (< comm. *etorri da*) 'has come', etc. As a matter of fact, this process has sometimes produced ultra-corrections

such as *idu* (< *hiru*) ‘three’, *aidean* (< *airean*) ‘in the air’ and *usadioa* (< *usarioa*) ‘the usage’, *amodio* (< *amorio*) ‘love’.

Although present in a certain number of loanwords, word-initial /d-/ is almost entirely absent from the native vocabulary; the only exception being represented by the auxiliary verbal system, where initial /d-/ is very common, contrasting with its scarce presence in lexical items. Alongside these common auxiliaries, some verbal forms in the subjunctive and conditional (*ledin*, *letorke*, *lioke*, etc.) have an initial *l-* which seems to correspond to the initial *d-* widely present in the indicative (*da* ‘is’, *dira* ‘are’, *dut* ‘I have’, etc.). As noted by Michelena (*FHV*: 257), some Latin and Romance borrowings also show an initial lateral instead of the expected voiced dental stop: *leka* (< **deka* < Lat. *thēca*) ‘pod’, *lanjer* (< Fr. *danger*) ‘danger’, *libersio* (cf. Sp. *diversión*) ‘diversion’, etc. These correspondences lead us to conclude that the initial /d-/ regularly became an alveolar lateral at some time within the Proto-Basque period, thus predating the development of the dialects. The maintenance of the initial /d-/ in the indicative may be a trace pointing towards the suspicion that such forms were preceded by more phonological material when that particular change happened.

Also word-initially, /b-/ has been lost before a back vowel. This process is almost systematic before /o/ —with *bost* ‘five’ being one of the few exceptions— and it is also usual before /u/: *otu* (< Lat. *uōtum*) ‘request’, *okela* (< Lat. *buccella*) ‘piece, bit, mod. meat’, *buztarina* > *uztarina* (< Lat. *postilēna*) ‘cushion to take baggage on the rump of a horse’, etc.

4.4.2.2. Voiceless stop series

Although the modern language shows no difference between the number of voiced and voiceless stops, older stages of the language lack the labial voiceless stop that would have been the counterpart of /b/, as was made apparent by Michelena (*FHV*: 261).

As a matter of fact, the labial voiceless stop /p/, which nowadays is as common as any other stop in the inventory, becomes more and more scarce the further we go back in time. The general absence of /p/ in patrimonial words —since most words containing /p/ may be analyzed either as loaned vocabulary or compounds— along with its widespread phono-symbolic use, make it very difficult to accept its presence in older stages of the language.

As mentioned above, the modern language has voiceless stops not only in onsets —both in initial and medial position— but also in some codas in word-final position. Some centuries ago, not only were the voiceless stops fewer than they are now in the modern dialects, but they were also more constrained as to the positions in which they could appear.

Voiceless stops were completely absent from the word-initial position, as evidenced by the adaptation of loanwords from Latin and Romance, in which the initial stops were systematically neutralized to voiced, as shown by the following examples: Lat. *pācem* > *bake* 'peace', Lat. *peccātum* > *bekatu* 'sin', Lat. *parcere* > *barkatu* 'forgive', Lat. *tempora* > *denbora* 'time', Sp. *torre* > *dorre* 'tower', Sp. *castaña* > *gaztaña* 'chestnut', Lat. *catēna* > *gatea* 'chain', etc. The instances of initial voiceless stops are either modern borrowings or result from a later assimilation, as in the evolution of the word *katea* < *gatea* < **gatena* < Lat. *catēna* 'chain' (see §4.4.5.1).

It is also worth mentioning that no intervocalic voicing comparable to that of peninsular Romance languages developed within Basque, so that all intervocalic voiceless stops borrowed from Latin remained voiceless. However, another neutralization can be observed in the adaptation of Latin and Romance word-medial stop clusters and geminates, which were systematically borrowed as voiceless stops. This occurred not only with the voiceless clusters, but also with the etymologically voiced clusters: Lat. *abbas* > *apaiz* 'priest', Old Sp. *cobdiçia* > *gutizia* 'desire, whim', Lat. *sabbatum* > *zapatu* 'Saturday', Lat. *ad uallem* > *apal* 'downward, humble', Lat. *captīvum* > *gatibu* 'captive, prisoner', etc. A possible approximant production of the intervocalic voiced stops from early times (see §4.4.2.1 above) may have conditioned the perception of foreign medial stop clusters as stops rather than approximants, which may have been the only other available option. The heterorganic stop clusters (both voiced and voiceless) have been regularly adapted as taking the place of articulation of the second segment. This is probably a consequence of stop place cues being better perceived in prevocalic than in postvocalic position (Ohala 1993).

Coda stops, which are even today restricted to a reduced number of morphemes, are the result of an apocope and the devoicing of the final voiced stop, and have been introduced into the language in recent times (cf. Artiagoitia 1993). Examples of these are the development of the morpheme *-k* (< *-ga*, still present in toponymy), currently used as ergative marker (cf. Lakarra 2006b) as well as for pluralizing in nominal morphology, and the morpheme *-t* used for the first person in verbal morphology (< *-da*, cf. *dut* 'I have' versus *dudan* 'that I have'). Michelena mentions that Oihenart wrote *dud* for the modern *dut* 'I have' in *Notitia utriusque Vasconiae, tum ibericae, tum aquitaniae*, in 1638 (*FHV*: 236).

The voicing opposition was also partially neutralized after /l/ or nasals, but it favored the voiced series in this context. However, both of the easternmost dialects, Roncalese and Souletin, have kept the distinction: Lat. *tempora* > *denbora* (but R. *tenpra*) 'time', comm. *igande* (but R. S. *igante*) 'Sunday', comm. *sendo* (but R. S. *sent(h)o*) 'strong', etc. Another neutralization occurred after fricative

sibilants, where voiced stops became voiceless, as in *ez du* > *eztu* ‘has not’, *ez da* > *ezta* ‘is not’, *deskana* (< Sp. *desgana*) ‘faint’, etc. This process was probably systematic at some point and is traceable to at least the Middle Ages, as shown by place names such as *Erdozpea* (cf. *behea*) from 1143 and *Çuastuy* (cf. *-dui*) from 1284.

4.4.2.3. Aspirated voiceless stop series

Alongside the voiced and voiceless stop series, modern eastern dialects also possess a series of aspirated voiceless stops composed of /p^{h}/, /t^{h}/, /k^{h}/ and the not so usual palatal /c^{h}/.[39] This series does not seem to have an autonomous status in older stages of the language, since the aspirated stops mainly behave as allophones of the voiceless series, whose presence was probably conditioned by prosodic prominence. An evidence of their secondary status is the development of such segments in the place of etymological voiceless stops in Latin and Romance loanwords, such as S. *phena* (cf. Lat. *poena*, Sp. *pena*) ‘sorrow’, S. *merkhatü* (cf. Lat. *mercātus*, Sp. *mercado*) ‘market’ and S. *apho* (cf. Sp. *sapo*) ‘toad’.

As a consequence of their prosodically conditioned allophonic status, their possible contexts are mostly limited to the onset of the first two syllables (cf. Michelena 1988 [1951a]), following the restrictions also met by aspiration (see §4.4.3.1.1 and §4.6.4). Further constraints applied to an aspirated stop are that it cannot appear in words containing an aspiration, and that it may only appear once per word, albeit this latter restriction is similar to the one applied to aspiration. When a word —especially a loanword— contains two different voiceless stops subject to becoming aspirated (i.e. in the onsets of the first two syllables), it is regularly the first syllable which is produced as the aspirated one, if any.[40] Examples of this restriction include *khorpitz* (< Lat. *corpus*) ‘body’, *phiper* ‘pepper’, *khatiña* (< Lat. *catēna*) ‘chain’ and examples like *tenta* ‘tempt’ with no aspiration, but structures like **TVT^{h}V and **T^{h}VT^{h}V —as well as **hVT^{h}V and **T^{h}VhV— are unattested.

Nevertheless, in the Souletin dialect, there are some words in which a stop may be produced with aspiration even after the second syllable, if it still precedes the stress —usually located in the penultimate syllable, but not necessarily

39 See §4.4.4.1 on palatal segments and their special status within the language.

40 This contrasts with aspiration, since /h/ is regularly kept on the second syllable instead of the first, which is usually dissimilated in words etymologically containing two aspirations (cf. §4.4.3.1.1).

(cf. §4.6.3)— as in *boronthate* ‘will’, *ba(r)anthalla* ‘february’, *arrathũ* [ara'tʰũ] ‘mouse’, etc.

4.4.3. Fricatives

4.4.3.1. Different kinds of aspiration

4.4.3.1.1. Voiceless glottal fricative /h/

Although nowadays it is only present in eastern dialects, some centuries ago aspiration was common to all Basque varieties. Michelena (*FHV*: 203) describes it as being voiceless in word-initial position but usually voiced between two vowels, and nasal between two nasalized vowels.

According to Michelena, its loss could have begun in Navarre in the 11th century, but it was still present in Alava and La Rioja until at least the 13th century (*FHV*: 205). It is already absent in the Biscayan texts from the 16th century. Its presence was diminishing in modern Labourdin and was already lost in its coastal variety,[41] at least from the 19th century, according to Bonaparte’s (1991 [1869]) descriptions. Its frequency is higher in Low Navarrese and at its highest in Souletin.

The aspiration seems to be limited to the (prevocalic) onset position (but see §4.4.1.1 about a hypothesized coda /h/), and it is absent from any kind of tautosyllabic cluster. It can appear in the following contexts: in initial position, between two vowels (or a diphthong and a vowel) and between a heterosyllabic sonorant (/n/, /l/, /ɾ/, /r/, /ɲ/ or /ʎ/, but never /m/ or a sibilant) and a vowel: *harri* ‘stone’, *haur* ‘child’, *ahizpa* ‘sister (of a woman)’, *oihan* ‘forest’, *senhar* ‘groom’, *üñhürri* ‘ant’, *elhe* ‘word’, *illhargi* ‘moon’, *erhi* ‘finger’, *urrhets* ‘step’, etc. Although the rhotics were neutralized in coda position, their contrast was preserved before an aspiration (apud *FHV*: 203). This was the case in Souletin until very recent times, when the opposition between rhotics was systematically lost in all environments after the loss of the alveolar tap (cf. §4.4.1.1).

A further limitation as to the location of the aspiration implies that it is restricted to the first two syllables of the word. As Michelena (1988 [1951b], 1988 [1957/58]) proposed, this may be due to a demarcative stress pattern, according to which every word was stressed in the second syllable from the left, hence the aspiration was limited to pre-stressed position (cf. §4.6.4). This constraint may have biased the metathesis of several secondary aspirations that fell from the

41 Although it was systematically used in written documents from the 17th century (see *FHV*: 204).

third syllable onwards (cf. §4.4.5.3), such as Lat. *arēna* > **areha* > modern *harea* ‘arena’ and Lat. *leōnem* > **leohe* > modern *lehoi* ‘lion’.

There is yet another restriction which affects the aspiration in the modern dialects, although it was not present in medieval times: no Basque word contains two aspirations or two aspirated segments (since aspirated stops are also constrained by this restriction, cf. §4.4.2.3). Words etymologically containing two aspirations have been simplified following a development similar to Grassmann’s Law of Greek and Indo-Iranian, i.e. an anticipative dissimilatory loss has affected every aspiration followed by another within the boundaries of the word (see §4.4.5.2).

A reduced number of word-initial aspirations were also introduced along with loanwords which had /f-/ in their Latin forms, such as *horma* (cf. Lat. *forma*) ‘wall’, *haro* (cf. Lat. *pharum*, Sp. *faro*) ‘lighthouse’, *haxe* (cf. Lat. *fascem*) ‘beam, bundle’ and *holla* (cf. Lat. *foliam*) ‘leaf, sheet’ (see §4.4.3.2). However, this development may have occurred before the words were borrowed, and thus within the Romance languages themselves.

There are also some instances of /h/ that appear in composition, which are derived from a former final tap in the first member of the compound in words, such as *uharte* ‘island’, *uhalde* ‘shore, riverside’ and *zuhandor* ‘shrub, bush’ (see §4.4.1.1 on these).

4.4.3.1.2. Nasalized /h̃/

Alongside the etymological /h/, we also find a secondary aspiration which derived intervocalically from an older alveolar nasal stop /n/ (cf. *FHV*: 300-303, Igartua 2011 and §4.4.1.3.1). This aspiration /h̃/ was historically nasalized and probably also voiced (hence it could be transcribed as /ɦ̃/), and this is still the case for the Souletin dialect. It has seemingly merged with the other glottal fricative in the remaining aspirating dialects, although a description of its voicing is yet to be produced.

As a consequence of its origin, the modern context of the /h̃/ is limited to the intervocalic position, where it contrasts with the non-nasalized aspiration /h/ in the Souletin dialect. It was most probably common to all dialects during a medieval stage of the language, since the loss of the intervocalic /n/ is regular in all of the varieties which no longer preserve the aspiration today.

Examples of this development are especially explicit in Latin borrowings such as Lat. *honōre* > L. *ohore*, S. *uh̃ure* ‘honor’, Lat. *anatem* > S. *ah̃ate* ‘duck’ and S. *ah̃abe* (cf. Sp. *anavia*) ‘blueberry’. This aspiration was also lost (§4.4.3.1.1) or metathesized when located after the second syllable, as shown by

Lat. *corōna* > *k(h)oroa* ‘crown’, Lat. *capitāne(m)* > Old B. *kapitãẽ* > modern *kapitain* ‘captain’, Lat. *arēna* > *(h)area*, R. *ãria* ‘sand’.

4.4.3.2. The emergence of the labio-dental fricative /f/

Although /f/ is mainly labio-dental in the modern Basque dialects, Hualde reports some bilabial realizations (*GB*: 15). The labial fricative is proposed to be absent in the proto-language (*FHV*: 262), and the fact that /f/ is mostly found in borrowings provides evidence to suggest that this segment is of recent introduction.

Alongside its limited presence in native vocabulary, /f/ seems to vary to a great extent with other labial segments, generally /p/ or /b/, which could later become a nasal /m/. For the sake of argument, in Guipuscoan, /p/ is the usual outcome of /f/ in borrowings such as *paltsu* (< Lat. *falsum*) ‘false’, *peria* (< Lat. *fēria*) ‘fair’ or *inpernu* (< Lat. *infernum*) ‘hell’. The development also works in reverse, with etymological voiceless labial stops becoming voiceless labial fricatives in examples such as *frogatu* (< Lat. *probātum*) ‘proved, to prove’, S. *foltsü* (cf. Lat. *pulsum*) ‘pulse’ or common *prakak* > B. *frakak* ‘trousers’.

The Romance bilabial voiced fricative <v> /β/[42] is usually adapted to /b/, which is also produced as [β̞] between vowels. Nevertheless, some words have the voiceless fricative /f/ instead of the expected voiced stop in initial position: LN. *ferde* (cf. Sp. *verde*) ‘green’, HN. *fau(n)* (cf. Sp. *vano*) ‘vain’, etc.

As well as labials, the Latin /f/ can also result in a Basque aspiration, albeit this may have been after Romance (Gascon or Spanish) mediation. In such borrowings, some dialects show a labial while others display an aspiration (in eastern dialects) or zero (in western dialects).[43] Examples of this alternation are: L. *haro* (cf. Sp. *faro*) ‘lighthouse’, *haxe* (cf. Lat. *fascem*) ‘beam, bundle’, *holla* (cf. Lat. *foliam*) ‘leaf, sheet’ in the eastern dialects; and in the western dialects, *piku(o)*, *biko*, *fiku(o)* : B. *iko*, HN. *igu* (cf. Lat. *fīcum*) ‘fig’ and *biru*, *piru*, *firu* : B. *iru* (cf. Lat. *fīlum*) ‘thread’.

Hualde has proposed that the word-medial /-f-/ arose within the native vocabulary from the combination of a labial —which could be a wau or an approximant [β̞]— and an aspiration (Hualde 1997a: 422-423). More specifically, Hualde suggests that *auhari* ‘dinner’ and *auher* ‘lazy, idle’ are the oldest variants in their dialectal cognate groups —cf. *afari*, *abari*, *apari*, *aufari*, *auhari*, *aihari*, *aigari*, etc., for the former; and *alfer*, *alper*, *aufer*, *aulfer*, *aurer*, *aulger*, *auher*, etc. for the latter— and that their medial clusters (i.e. *-βh-* or *-wh-*) would

42 See Penny (2006: 118-119).

43 Remember that /h/ was lost in western varieties (cf. §4.4.3.1.1).

yield /f/ in some dialects (cf. *afari*) and /b/ (as in *abari*) wherever the aspiration was already lost (cf. §4.4.3.1.1). While his hypothesis seems rather plausible, the proposal of an assimilation may be preferable to that of coalescence, since this would account for the variants which kept the /w/ (cf. *au̯fari* and *au̯lfer*). Moreover, if the /l/ in *alfer* is etymological —which would appropriately relate the forms *abari* and *aurer* as results of an early loss of the aspiration— it could also permit the presence of an intervening segment such as that of **au̯lher* > *au̯lfer*, *alfer*.

4.4.3.3. Sibilants

4.4.3.3.1. Voiceless fricative and affricate sibilants

Until recent times, all Basque varieties have opposed a series of apico-alveolar voiceless sibilants to a lamino-alveolar series. These series consist of a fricative (/s̻/ or /s̺/) and an affricate sibilant (/ʦ̻/ or /ʦ̺/), which contrast in certain positions —e.g. intervocalically— but are (almost) absent from others. A series of the prepalatal sibilants /ʃ/ and /ʧ/ is also present in all dialects, and, although they have historically been used as a device to develop affective variants (cf. §4.4.4.1), they are common in all modern dialects.

In most dialects, /ʦ̻/ and /ʦ̺/ are absent from word-initial position, where only their fricative counterparts occur. They only appear in a few expressive words such as *tzar* ‘evil’ and *tzakur* ‘big dog’ and only in certain varieties (Hualde *GB*: 22). The distribution of the more modern prepalatal sibilants is the opposite, with only the initial /ʧ/ present in most dialects, although both /ʃ-/ and /ʧ-/ occur in Souletin and only the fricative occurs in Labourdin.

Word-medially, all of the sibilants contrast intervocalically, while only affricates appear after a nasal or a liquid in prevocalic position, and only fricatives appear after a stop or before a consonant: *hotz* ‘cold’ but *hoztu* ‘cool, chill’ or *beltz* ‘black’ but *belztu* ‘blacken’. However, the stop/sibilant clusters are very modern and are usually found in very recent borrowings such as *akzio* ‘deed’ or *seksu* ‘sex’. The absence of fricatives after liquids or nasals is a consequence of the affrication of sibilants in this context and the subsequent contextual neutralization of the opposition it produced. This is appreciable in the modern Basque forms of Romance borrowings such as *pultso* (cf. Sp. *pulso*) ‘pulse’, *bertso* (cf. Sp. *verso*) ‘verse, line’ and *kontzientzia* (cf. Sp. *conciencia*) ‘conscience’. This development was not thoroughly completed after rhotics, but it is systematic after both nasals and laterals.

In word-final position, affricates are much more frequent than fricatives, which although still present, are significantly reduced in number. This is the re-

sult of an unfinished affrication process —similar to that which occurred after liquids and nasals— which has affected most final sibilants but which has not become systematic. In any case, the Latin final /-s/ has regularly affricated when adapted to Basque: Lat. *fortis* > *bortitz* 'strong', Lat. *corpus* > *gorputz* 'body', Lat. *opus* > *oputz* 'physical effort', etc.[44]

Insofar as loanword adaptation is concerned, the Latin /s/ has regularly been borrowed as the lamino-alveolar sibilant /s̻/: *sigillum* > *zigilu* '(postage) stamp', *sellam* > *zela* 'saddle', *paradisum* > *baradizu* 'paradise', *spatam* > *ezpata* 'sword'. Only more recent borrowings such as *musika* 'music' or *soinu* 'sound' have an apical corresponding to the Romance alveolar sibilant, and only in the western dialects that are in contact with Spanish, since French sibilants are still borrowed with /ʂ/ (cf. Fr. *autobus* > *otobuz* 'bus'). In modern borrowings, the Spanish interdental fricative /θ/ is regularly adapted to the lamino-alveolar fricative, which may also be affricate in a limited number of examples: Sp. *plaza* [plaθa] > *plaza* [plaʂa] 'square', *hacienda* [aθjen̪da] > *hazienda* [aʂjen̪da] 'Treasury, public funds', *pozo* > *putzu* 'well'.

In the modern Biscayan dialect, as well as in some Guipuscoan varieties, the two alveolar series of sibilants have merged so that every fricative is produced as an apical and every affricate as a laminal sibilant. This process was already complete in western Biscayan texts from the 17th century[45] onwards, and has gradually expanded westwards into Guipuscoa, first spreading through larger towns and later through the villages, so that it was already completed in Donostia-San Sebastián by the first half of the 18th century. In other dialects, sibilant fricatives have been biased towards the apical in certain contexts, although not exhaustively. Such contexts include preconsonantal and final positions (cf. *FHV*: 282).

4.4.3.3.2. Voiced alveolar sibilants

As in Spanish, voiced sibilants are only allophones of their voiceless counterparts in most Basque dialects, and they are produced when followed by a voiced consonant, as in *asmatu* [aʐmatu] 'invent, guess correctly', *esnatu* [eʐnatu] 'wake up' and *ezdeus* [eʐð̞eu̯ʂ] 'nothing'. However, modern Souletin and some Low Navarrese areas possess voiced alveolar sibilants (/ʐ/, /ʐ̻/ and /ʤ̻/) as phonemes, although they mainly appear in recent borrowings or as shandi effects: S. *arrazũ* [araʐ̻ũ] 'reason', *plazent* [plaʐ̻en̪t] 'pleasant', *etsenplü* [eʤ̻emply] 'example' and *deüs ere* > *deüse* [deyʐe] 'something, nothing (when negated)'.

---

44 There are, however, rare exceptions, as shown by Lat. *magis* > *maiz* 'often'.

45 Michelena mentions Micoleta and *Viva Jesus* (*FHV*: 282).

4.4.4. Palatal series

4.4.4.1. Palatal segments and their usage

An old series of palatal segments which is found in all of the modern dialects, contrasts with the non-palatalized consonants. It can be traced back to older stages of the language, but its use was probably marginal and banned from the standard inventory. They were almost certainly restricted to the creation of hypocoristic or affective variants, which were formed without derivative additions; as well as being present in the borrowed lexicon, e.g. *zikoña* 'stork' from Romance (cf. Lat. *cicōnia*). However, these segments have not themselves proven to be productive insofar as historical reconstruction is concerned, and most authors have systematically worked from their non-palatalized equivalents. The series of palatal segments is usually formed by /ɲ/, /ʎ/, /ʃ/, /ʧ/, /c/ and /ɟ/, along with the segments derived from /j/, which are dialectally divergent.

Although palatals were usually used as a morphological device, they could also arise from assimilation due to their phonological context, or even become incorporated through Romance borrowings. In such cases, they would have resulted from contact with a preceding palatal segment, thus being practically absent from word-initial position. Nevertheless, note that hypocoristic palatalization can occur in any given context, which means that affective variants are the only words with an initial palatal in all of the dialects in which the yod has become a non-palatal segment.[46]

Given that the creation of palatal from non-palatal segments in words with affective functions was undoubtedly sporadic, this section will focus on the palatalizations linked to phonological reasons.

4.4.4.2. Assimilations

Assimilation to a previous palatal segment (an /i/, a palatal off-glide or even a fronted consonant in very specific contexts) is regular in many varieties of Basque. There are a handful of segments which may undergo this kind of palatalization.

The sonorants are the consonants whose palatalization is most dialectally widespread. /ɲ/ and /ʎ/ appear regularly after a palatal segment in many Basque varieties: *iñor* 'nobody', *ora(i)ñ* 'now', *ba(i)ña* 'but', *illargi* 'moon' and *ille* 'hair'. /ɾ/ and /r/ are also palatalized to /ʎ/ in eastern dialects: *bero* > *bello* 'hot', *hori* > *holli* 'yellow', *hurran* > *hüllan* 'near'. This process mainly takes place after a palatal glide, but is also usual after a syllabic /i/. When an off-glide is in-

46 See §4.3.4.3 on this.

volved, the assimilation may result in an apparent coalescence with the following segment.

Less frequent, but also usual, is the palatalization of sibilants. This does not result in a "standard" palatal but rather a palato-alveolar /ʃ/ or its affricate counterpart /ʧ/. This is particularly common in Biscayan, and occurs most frequently with the etymologically laminal sibilant:[47] *(h)aize* > *axe* 'wind', *ixan* 'to be', *itsu* > *itxu* 'blind', *bakoitz* > *bakotx* 'each one', etc.

In some areas,[48] /t/ is palatalized to /c/ under the same circumstances (as in *aitta* 'father', *zaittut* 'I have (...) you' or *kitto* 'enough, even'), while in some Biscayan and Guipuscoan areas it has undergone further changes to /ʧ/ (*aitxa*, *zaitxut*, etc.).

The palatalization of /d/ to /ɟ/ (contextually spirantized to [ʝ̞] as in the rest of the voiced stop series) is not as extended as the aforementioned examples, but it is still present in some dialects. Guipuscoan and High Navarrese show examples of this: *fidatu* > *fiyatu* [fiʝ̞atu] 'confident, trusting', *gidatu* > *giyatu* [giʝ̞atu] 'guided'. It even appears after a palatal nasal or lateral (as in *iñddar* [iɲɟar] 'strength' and *billddur* [biʎɟur] 'fear'). There are also some instances of palatalization of the voiced velar obstruent: *igo* > *iyo* [iʝ̞o] 'go up'. This segment can be produced both as a voiced palatal stop /ɟ/ and as a voiced palatal fricative /ʝ/, and it can even yield an affricate /ʤ/ in some Biscayan areas.

Insofar as affective palatalizations are concerned, these stops —not only the dental, but also the labial or velars— are usually palatalized to the more habitual affricate /ʧ/ instead of the less usual palatal stops, although the latter are also dialectally present.

4.4.4.3. Outcomes of **j-*

The Basque yod (cf. §4.3.5.2) has developed a wide range of dialectal variants, some being more extended than others. These changes, which involve (pre-)palatal or even velar outcomes, appear to be diverging results along the same lines of evolution. They are illustrated in Figure 4.1.

Although the first group appears to have maintained the oldest state, the third may be the most conservative dialect, at least if it reflects the segment that was common to all Basque dialects, which may have been a fricative /ʝ/, and whose evolution may have developed a wide range of fricatives, and even affricates.

---

47 Nowadays merged with the apical in this particular dialect, cf. §4.4.3.3.1.

48 Only in Spanish territory, according to Michelena (*FHV*: 185).

Segment	Proposed development	Approximate dialectal extension (apud *FHV*: 168-171)
[j]	j (> ʝ > j)[49]	North.-HN, most L. and some B. areas
[ʒ]	j > ʝ > ʒ	S, some northwestern B. areas
[ʝ]	j > ʝ	Some B., LN. and HN. areas
[ɟ]	j > ʝ > ɟ[50]	LN
[ʤ]	j > ʝ > ʒ > ʤ	Northwestern B
[ʃ]	j > ʝ > ʒ > ʃ	Aesc, Sal, R, most South.-HN. areas
[x]	j > ʝ > ʒ > ʃ > x	G, northwestern HN, eastern B

Figure 4.1. Dialectal evolutions of the yod

As Michelena (*FHV*: 170) points out, the parallel evolution of Romance palato-alveolar fricatives appears to be crucial for the dialectal distribution of the modern Basque fricatives, since only the varieties in contact with Spanish have developed voiceless fricatives from **j-*.

4.4.4.4. Depalatalization

Depalatalization is common in Low Navarrese, Labourdin, and is even regular in the High Navarrese of Sakana. Where other varieties show /ɲ/ or /ʎ/, these dialects have a palatal off-glide in the previous syllable. This even occurs with the etymological palatals, as in *boteila*, *botila* 'bottle' from Sp. *botella* /boteʎa/.

4.4.5. Some consonantal processes

There are some processes which do not affect particular segments but rather specific phonological features, or even complex groups of features. Many of these deserve a special overview, but here we will focus on the analysis of three general kinds of processes: assimilations, dissimilations and metatheses.

4.4.5.1. Assimilation

Alongside the local assimilation of the palatality mentioned in §4.4.4.2, there are some other processes of local assimilation within the Basque language, although

49 At least for the Biscayan area, a round-trip development is reconstructed by Hualde & Elordieta, G. & Elordieta, A. (1994: 8-9).

50 Produced as an approximant in most contexts (§4.4.2.1).

they are by no means as widespread as the one produced by front high vowels. The most important among them is the local assimilation of labiality briefly mentioned in §4.4.1.3.1, which is also perseverative. In this process, a preceding /u/ labializes the following consonant, as in *un(h)il* > L. *umil* 'funnel' or **zunai* > HN. *zumai* 'forage', etc.

There are three main processes of distant consonantal assimilation within the history of the language, namely an assimilation of the nasal feature (cf. §4.4.1.3.2), an assimilatory process that occurs among stops (cf. §4.4.2.2), and an assimilation affecting the place of articulation of sibilants.

The first process involves the nasalization of a stop (usually labial) when followed by a nasal obstruent, and has been very productive, to the extent of greatly increasing the frequency of the labial nasal, previously absent, in the language.[51] This assimilation is particularly productive when only a vowel separates the segments involved in the process, which do not necessarily need to be tautosyllabic, as in *heben* > *hemen* 'here', *bendekatu* > *mendekatu* (cf. Lat. *uindicāre*) 'avenge', *menia* (< Lat. *uenia*) 'truce' and *bañu* (cf. Sp. *baño*) 'bath'. Along with these examples, there are also instances of such nasalization when there is a greater gap between the participating segments, as in *bagiña* > *magi(ñ)a* (cf. Lat. *uāgīna*) 'scabbard, sheath' and *baemana* > *maemana* 'already given'. Nevertheless, there seem to be certain environments in which this assimilation is blocked, since sequences with an intervening coda rhotic, such as *bVr.n*, never become ***mVr.n* (cf. Egurtzegi 2011, §4.4.1.3.2).

The second process involves the devoicing of a stop when followed by a voiceless stop in the next syllable. This process has been a very common mechanism for the development of initial voiceless stops in both the patrimonial and the old loaned vocabulary, since some stages of Proto-Basque lacked voiceless stops in word-initial position, and thus early Latin loanwords were not borrowed with word-initial voiceless stops (see §4.4.2.2). Examples of this process are *bihi gain* > **bit-gain* > *bikain* > *pikain* 'best, first', *behi gorotz* > **bet-gorotz* > *pekorotz* 'cow dung', **gurd*(*i*)*-bil* > *gurpil* > *kurpil* 'wheel' *gal-te* > *kalte* 'harm, damage' (cf. *galdu* 'lose'), Lat. *pācem* > *bake* > *pake* 'peace', Lat. *catēna* > *gat(h)ea* > *kat(h)ea* 'chain', *putre* (cf. Sp. *buitre*) 'vulture', etc.

The final type of assimilation can be inferred from the high frequency of Basque words involving two homorganic sibilants, as noted by Michelena (*FHV*: 283), who explained most exceptions by analogy (e.g. with suffixes like *-tze*). Examples of this homorganicity include: *isats* 'tail', *itsaso* 'sea', *itsatsi* 'stick, glue', *itsusi* 'ugly', *sustrai* 'root', *zazpi* 'seven', *zezen* 'bull', *zorrotz* 'sharp', *zortzi* 'eight', *izotz* 'ice', etc; and examples of place assimilations are

51 Along with other processes and its presence in the borrowed vocabulary, see §4.4.1.3.2.

frantzes > *frantses* 'French' (cf. *Frantzia* 'France'), *sazoi* > *sasoi* (cf. Sp. *sazón*) 'seasoning' and *zin(h)etsi* > *sin(h)etsi* 'believe', which was still written with an initial laminal by Dechepare in the 16th century.

4.4.5.2. Dissimilation

Dissimilation occurs when a non-etymological reinterpretation results from an ambiguous sequence, including more than one segment with a particular phonological feature which contains an elongated phonetic cue (apud Ohala 1981 and 1993). It can affect either that particular feature (removing it from one of the segments which bears it) or the whole segment which possesses it (by removing the affected segment from the sequence). Segments affected by dissimilation include, among other phonological features, those bearing rhoticity, laterality, nasality, labiality and aspiration. However, the only systematic process of total dissimilation found in Basque takes place in a word which contains two aspirations.

In this process, usually known as Grassmann's Law in the Indo-European tradition, a sequence of two aspirations is systematically simplified by deleting the first of them. In Basque, it has affected both the oral /h/ and the nasalized aspiration /h̃/. Examples of this kind of dissimilation are mainly found in compounds such as *hil* + *hotz* > *ilhotz* 'corpse', *hil* + *herri* > *ilherri* 'graveyard', *hil* + *hargi* > *ilhargi* 'moon', etc.

We find further examples of dissimilations, which, as stated above, have only been sporadic. These include words such as *ardore* > *adore* 'fortitude, courage'; **artxeter* > *atxeter* (cf. Lat. *archiater*) 'physician'; *errierta* > *errieta* (cf. Sp. *reyerta*) 'quarrel', with the loss of a coda rhotic; *alkar*, *arkal* < **hark-har* 'mutually'; *berar* > *bedar*, *belar* 'grass'; **berarri* > *belarri* 'ear'; *hortzadar* > *holtzadar* 'rainbow', with loss of the rhoticity; **eskinantzak* > *eskidantzak* (< Bearn. *esquinances*) 'tonsils'; **izenonbreak* > *izelonbreak* 'name and surname'; *puldamentu* (< Sp. *fundamento*) 'foundation'; with loss of the nasality; *baberruma* > *baberruna* 'bean', for labiality, etc.

4.4.5.3. Metatheses: reciprocal metathesis, perceptual metathesis and repercussion

There are two different kinds of metathesis which have been very productive throughout the history of the language (cf. Egurtzegi 2011), although neither of them seems to have been systematic in any particular context.

The first, namely reciprocal metathesis, affects two articulatorily similar segments located in the same syllabic position (i.e. both either on syllable onsets

or codas) in nearby syllables (usually consecutive). Some examples of these are the metathesis involving two coda liquids, as in *alper* : *arpel* 'lazy', *ergel* : *elger* 'stupid' and *sorbalda* : *solbarda* 'back (of the body)'; the metathesis of an onset rhotic and an onset approximant: *irago* : *igaro* 'go past', *irudi* : *iduri* 'image', *edarra* < *errada* (< Sp. *herrada*) 'wooden bucket' and *madarikatu* < *maradikatu* (< Lat. *maledīcere*) 'damn, curse'; the metathesis between an onset nasal and an onset lateral: *belena* (cf. Fr. *venelle*, Old Navarrese *benela*) 'gap between two houses', *il(l)unabar* > *iñulabar* 'dusk', *labana* < *nabala* (cf. Sp. *navaja*) 'knife'; and the metathesis of two onset obstruents agreeing in continuancy: *bage* > *gabe* 'without', *bilgor* > *gilbor* 'belly, paunch', *bekatari* > *betakari* 'sinner', *eskapatu* > *espakatu* 'escape, get away', *bezatu* : *zebatu* 'gotten used to', *ezabatu* (< Lat. *effaciāre*, cf. Occ. *esfaçar*) 'erase', *erakutsi* > *eratsuki* 'show' and *petxatu* (< Sp. *petacho*) 'patch'.

The second type of metathesis is directly related to dissimilation, as understood by Ohala (1981, 1993). As previously mentioned, features with elongated phonetic cues can create ambiguity in the string, which in some cases may be reinterpreted by the listener, thus yielding metathesis (Blevins & Garrett: 1998 and 2004), or even repercussion.

The features (and subsequent segments) susceptible to being metathesized are similar to those involved in dissimilation, and the instances of this process are also sporadic: *gurtze*, *kurtze* (< Lat. *crucem*) 'cross', *estrata* > *estarta* (< Lat. *strāta*) 'path, track' and *kruel* > *kurel* (< Sp. *cruel*) 'cruel', with metathesis of the rhoticity; *estalpe*, *estalbe* (cf. *establia*, < Sp. *establo*) 'stable' and *euli* > *ebli* > *elbi* 'fly' for laterality; and *harea* (< **areñ̃a* < Lat. *arēna*) 'sand', *heuskara* (< **enuskara*) 'Basque' and *onherran* (< **honerran*) 'blessing' are examples of metathesis of the aspiration. Glides can also become relocated, as in *euskara* > *eskuara* 'Basque', *bakoitz* > *baikotz* 'each' and *hausin* > *hasuin* 'nettle'.

The case of repercussion (or copying of a segment) is similar to that of perceptual metathesis, but instead of reinterpreting the ambiguous consonant in a different location, the etymological position is maintained and a new segment is added in a different position. Examples of this process are: *altzairu* < *altzailu* < **atzailu* 'steel', *aitzina* > *aintzina* '(to the) front', *zentzu* > *zentzun* 'sense, consciousness'.

There are also some instances of feature metathesis, namely those of nasality and palatality: *holiñe* (< Sp. *hollín*) 'soot', *laño* (< Sp. *llano*) 'modest, affable', *señale* > *senalle* (< Sp. *señal*) 'signal', **artzani* > *artzãi* > *artzain* 'shepherd', **arrani* > *arrãi* > *arrain* 'fish', **lukanika* > *lukãika* > *lukainka* (cf. Vulgar Lat. *lucanica*) 'spicy pork sausage', Lat. *arēna* > **areã* > *ãrea* 'sand'.

4.5. Phonotactics

Given that there was a time when the language only had monosyllabic CVC roots (Lakarra 1995, etc.), every cluster must necessarily be considered secondary, both for consonants and vowels. However, various clustering and simplification processes must have played an important role in developing the modern language, in which certain clusters, but not any possible combination, are found in each word position.

Nuclei are always vocalic, since Basque lacks syllabic consonants.

4.5.1. Word and syllable initial positions

Not every segment was present word-initially in the Michelenian Proto-Basque. We reconstruct three voiced stops /b-/, /d-/ and /g-/, or, perhaps more precisely, a lenis stop series; /ʛ-/, /ʂ-/, /n-/ and /h/. This series of stops would later become voiced, as proposed by Martinet (1974 [1950]) and accepted by Michelena (*FHV*: 373-374). Initial /l-/ would later emerge instead of the coronal lenis stop /d-/ (as in *leka* < **deka* < Lat. *thēca* 'pod', cf. §4.4.2.1), which was thereafter systematically lost from this position. Old Latin and Romance borrowings were, according to Martinet's theory, systematically adapted with initial voiced stops, given that Basque probably neutralized both series in favor of the voiced one word-initially, or even lacked a second series altogether in this position: *baradizu* < Lat. *paradīsum* 'paradise', *gurutze* < Lat. *crucem* 'cross', *leun* < **deun* < Occ. *teun(e)* 'soft' (cf. however, the seemingly native *ke* 'smoke' and *kirats* 'stink').

Historically, there has never been any kind of cluster either in word initial or syllable-initial position until very recent times, when *muta cum liquida* types of clusters were not only allowed, but also produced. This is shown by examples such as the Romance loanwords *treku* < Sp. *terco* 'stubborn', *truku* < Sp. *turco* 'Turk' and *alpraketa* < Sp. *alpargata* 'espadrille', in which such clusters are produced by metathesis, or *andre* < *andere* 'lady', *eskla* < *ezkila* 'bell' and *mistrio* < Sp. *misterio* 'mystery', in which they are formed after a syncope.

In old borrowings, stop-liquid clusters were adapted by either of the following: anaptyxis, metathesis or deletion of the initial obstruent. Anaptyxis occurred mostly (although not exclusively) in clusters involving the rhotic, while deletion affected only obstruents which preceded a tautosyllabic lateral. We find instances of svarabhakti vowels, which are systematically copied from those following the cluster, in *liburu* < Lat. *librum* 'book', *gurutze* < Lat. *crucem* 'cross', *ferekatu* < Lat. *fricāre* 'caress, scrub', *garau(n)* < Lat. *grānum* 'grain' and *lukuru* <

Lat. *lucrum* ‘interest, greed’, and obstruent deletion in *laket* < Lat. *placet* ‘like’, *luma* < Lat. *plūma* ‘feather’, *loria* < Lat. *glōria* ‘glory’, *lore* < Lat. *flōrem* ‘flower’ and *eleiza* < Lat. *ecclēsia* ‘church’. Finally, metathesis could affect clusters with either the rhotic or the lateral, and seems to be more dialectal and much less systematic than the other processes described above: B. *kurtze* < Lat. *crucem* ‘cross’, Al. *kurel* < Sp. *cruel* ‘cruel’, *Gregorio* > *Gergorio* ‘Gregory’, *estalbe*, *estalpe* < Sp. *establo* ‘stable, barn’ (but cf. also Rom. *estalvi*) and HN. G. *kolka* < *kloka* (cf. Sp. *clueca* < *clocar*) ‘broody hen’.

See also Lakarra (2010a) on initial position and on the reconstruction of the left edge of the word.

4.5.2. Final position and coda tautosyllabic clusters

Only certain segments occur word-finally: sibilants, /n/, /r/ and less frequently, /l/. There are some instances of final stops in morphemes which can be found in both the declination and conjugation of the modern language, such as the ergative /-k/ and 1st person singular /-t/. However, these are secondarily formed from older voiced stop-vowel sequences (cf. *FHV*: 343 and Artiagoitia 1993, or §4.4.2.2), whose final stops lost voice after an apocope, as in **duda* > *dut* ‘I have’ or **niga* > *nik* ‘I (erg.)’ (cf. Lakarra 2006b). Certain modern dialects also contrast alveolar and palatal nasals and laterals in final position, but most do not. It is also possible that the Roncalese contrasted between rhotics word-finally (cf. Michelena 1988 [1953b], 1988 [1954a], Hualde *GB*: 34).

Final clusters, although probably secondary, are not rare in the modern language. These are formed by a sonorant /r/, /l/ or /n/, followed by a voiceless stop or an affricate, or a fricative sibilant followed by a voiceless stop. We find final clusters in *bost*, *bortz* ‘five’, *hartz* ‘bear’, *beltz* ‘black’, *ihintz*[52] ‘dew’, *bart* (< *barda*) ‘last night’, *galant* ‘gallant’. Final stops and clusters are also usual in examples of phono-symbolism and onomatopoeia, as in *zart* ‘smack’, *urrup* ‘slurp’ and *zaunk* ‘bark’.

4.5.3. Heterosyllabic clusters and medial position

The word-internal position is the only one in which most segments occur, since in the polysyllabic Michelenian Proto-Basque some segments, namely voiceless stops or taps, are only reconstructed between vowels.

52 This final laminal affricate has been proposed by Lakarra to be a suffix (1998a, 2003, etc.).

Medial clusters conserved in the modern dialects are usually composed of either a fricative sibilant or a sonorant (/r/, /l/ or a nasal) and a stop; or a sonorant and an affricate. These include: *asto* 'donkey', *oskol* 'shell', *ezti* 'honey', *azpi* 'down(side)', *erdi* 'half', *arto* 'corn', *argal* 'thin', *alde* 'side', *kalte* 'damage', *handi* 'big', *hanka* 'leg', *hertsi* 'close', *artzain* 'shepherd', *altxatu* 'stand up', *altzairu* 'steel', *intxaur* 'walnut' and *antzina* 'formerly'. Less frequently, we find clusters which contain a rhotic and either a lateral or a nasal, as in *erle* 'bee', *arnasa* 'breath' and *armiarma* 'spider'. Clusters containing a sibilant and a nasal are rare but attested: *asmatu* 'guess', *esne* 'milk' and *esnatu* 'wake up'.

However, modern clusters have not necessarily remained unchanged and, as mentioned in §4.4.2, most stops have spontaneously become voiced after a lateral or a nasal.[53] Clusters formed by a sibilant and a stop necessarily involve a fricative sibilant and a voiceless stop, since stops were devoiced after sibilants.

Although the aforementioned clusters have remained conserved in the modern dialects, most consonantal groups have been simplified, sometimes (but not necessarily) in a systematic way.

For clusters of two stops, these have been consistently adapted to a single voiceless stop which shares its place of articulation with the etymologically prevocalic stop (see §4.4.2.2), as in the following compounds: *begi* 'eye': **bet-gain* > *bekain* and **bet-buru* > *bepuru* 'eyebrow', **bet-belar* > *bepelar* 'eyelash', etc. Michelena (*FHV*: 346) suggests that a nasal in the second syllable may have had the same effect in some words, but only offers a couple of examples: *oroipen* < **oroit-men* 'memory' and *zipotz* 'peg used to close a barrel', composed (as he himself proposed) of *zihi* 'wedge, little stick' (> *zit-*) and *motz* 'short'.

Clusters formed by a stop and a following prevocalic sibilant have usually yielded an affricate in composition: *artzain* < **ard(i)-zani* 'shepherd', *betzain* < **beh(i)-zani* 'cowherd', *betzain* < **beg(i)-zain* 'inner membrane of the eye', *otsein* < **og(i)-seni* 'servant', *betsein* < **beg(i)-seni* 'pupil' (apud *FHV*: 346). The examples above may only correspond to clusters involving a coronal stop, since all stops —and even the aspiration— became /t/ in the final position of the first member of a compound (cf. *beg(i)* 'eye' + *ile* 'hair' > *betile* 'eyelash', etc.). However, this process also affected loanwords which included clusters with a non-coronal stop followed by a sibilant, as in *atsolutu* (cf. Lat. *absolūtum*) 'absolute', *etsamina* (cf. Lat. *exāmen*) 'exam', *etsenplu* (cf. Lat. *exemplum*) 'example'.

According to Hualde (*GB*: 23), in certain dialects the affrication of heterosyllabic stop+sibilant clusters can also occur across morpheme or even word

53 A neutralization developed in most dialects, with the exception of Roncalese and Souletin.

boundaries, across words with a high degree of syntactic cohesion: *bait zuen* > *baitzuen* 'because she/he had it', *onek zuen* > (Baztan) *onetzuen* 'this one had it'.

Two sibilants —either fricative or affricate— have regularly resulted in an affricate (*FHV*: 350-351). In the case of both sibilants being homorganic, the resulting affricate would also share their place of articulation (*ez-zen* > *etzen* 'it was not', *ez-zara* > *etzara* 'you are not', etc.), while a cluster formed by an apical followed by a laminal yields an apical affricate (**diots-zut* > *diotsut* 'I tell you', **ekus-zu* > *akutsu* 'see (you)', **sinhets-zazu* > *sinetsu* 'believe (you, imp.)', **hauts-tzaile* > *hautsaile* 'breaker'. However, Michelena (*FHV*: 351) finds no clear examples of two heterosyllabic sibilants yielding a laminal, but instead one which suggests the opposite: S. *egütseme* < **eguz-seme* 'godson' (cf. *egüzaita* 'godfather'); cf. also the etymologies **hortz-so* > *otso* 'wolf' and **hartz-so* > *atso* 'old woman', proposed by Lakarra. Any cluster involving a prepalatal sibilant would have a prepalatal outcome, regardless of its etymological position: **deretx-zun* > B. *deretxun* 'that you think', but also *gantxingor* < **gantz-txingor* 'crackling'.

Latin had some /mn/ nasal clusters that have been simplified in Basque. This simplification yielded either of the nasals involved: Lat. *damnum* > *damu* 'regret, damage', Lat. *autumnum* > B. *autono* 'autumn', Lat. *columna* > *koloma* 'column' or Lat. *domine* (> *dom'ne*) > *(jaun)done*, used with saint names.

Sequences of a nasal followed by a stop have sometimes been reduced to a single nasal, probably by assimilating the stop to the nasal (cf. **umbe* > *umme* —found in Lerga— > *ume* 'child' or *sembe* in Aquitanian anthroponyms > **semme* > *seme* 'son'). This process frequently affects labials, but is also found in alveolar clusters: *konbeni* > *komeni* 'to be appropriate, to be advisable', *konbentu* > *komentu* 'convent', *amizione* (cf. Lat. *ambitiōnem*) 'ambition', *gomitatu* (< Lat. *convitāre*) 'invite', *ganora*, *khanore* (< Lat. *candōrem*) 'skillfulness'.

The group of a sibilant followed by a sonorant —a cluster almost nonexistent in the modern language— has been simplified to a sonorant: *enaiz* < *ez naiz* 'I'm not', *igelenean* < **igaz-lenean* 'last year', *maialen* < *maiatz le(h)en* 'First of May', *mailasto* < *maiz-lasto* 'corn straw', *iraleku* < **iratz(e)-leku* 'bracken', etc. Michelena (*FHV*: 351) proposes that the coda sibilant would have become a voiced aspiration before being dropped.[54]

The clusters formed by a prevocalic lateral preceded by a sonorant have been simplified by deleting the first of the sonorants: *zala*, *zela* < *zan*, *zen* (plus the complementizer *-la*) 'that was (him)', *hola* < *hon(e)la* 'in this way, thus', B. *imillau(re)n* < **imin-lauren* 'measure for grain', B. *biloba* < *birloba* 'grand-

54 As mentioned in §4.4.3.3.2, sibilants are realized as voiced before a voiced consonant. Cf. also /-r/ > /-h/ in the first member of a compound, in §4.4.1.1.

child', B. *bialeku* < *bear-leku* 'workplace', or the place name *Goliz* (< *Gorliz*) in RS. However, /rl/ clusters are not absent from the modern language (cf. *arlo* 'task', *erle* 'bee', *gerla* 'war', *irla* 'island', etc.).

In consonantal clusters formed by three segments complex codas are simplified, while *muta cum liquida* onsets are much more common. In the case of a coda composed of a sonorant followed by a sibilant, the former is usually dropped: *belztu* > *beztu* 'blacken', *a(h)aztu* < *a(h)anztu* 'forget', *jazten* < *janzten* 'dressing', *ostu* < *onstu* 'stolen', *(h)osto* < *(h)orsto* 'leaf', *estu* < *herstu* 'tight', *ozpin* < *orzpin* 'lightning' and *asto* < *arsto* 'donkey'. The other case of complex coda described in §4.5.2 —formed by either a sonorant or a sibilant and a stop— would result in the stop at the end of the syllable being systematically lost, but only after devoicing the following stop (cf. the treatment of stop clusters supra): *ausarki* < *ausartki* 'bravely', *arkume* < *ard(i)-kume* 'lamb', *gurpegi* < *gurd(i)-begi* 'bucket' or *gurpil* < *gurd(i)-bil* 'wheel'. This intermediate stop is simply dropped before a *sonorant*, as in *izerleka* (from *izerd(i)* and *leka*) 'sweat that creates dirt' or *ipurñauska* (cf. *ipurdi*) 'jump, gambol'.

4.6. Accentuation

Modern Basque dialects have different types of accentuation, based on either stress or on pitch accent. However, there is a central area, including most of the High Navarrese, Low Navarrese and Labourdin varieties, which lacks any contrastive accentuation. In most High Navarrese varieties, stress automatically falls on the penultimate syllable, and it lacks any other accentuation pattern to contrast with, unlike Souletin, which has oxytonic words in addition to paroxytonic (cf. §4.6.3). According to Gavel (1920: 108), the Labourdin and Low Navarrese areas have no word-level stress (cf. Hualde 1993a: 31). Nevertheless, earlier types of accentuation may recently have been lost in the area (cf. Michelena 1987 [1972c]).[55]

The areas with distinctive accentuation surround this central territory and are located in both the western and eastern sides of the Basque Country (Hualde 1993a: 13). The most widespread accentual system is the central-western type, with peninitial stress, i.e., stress on the second syllable of the word.

55 Cf. also the Goizueta variety of High Navarrese (Hualde & Lujanbio 2008, and Hualde & Lujanbio & Torreira 2008), which still maintains a pitch accent system similar to the one found in Northern Biscayan (§4.6.2).

4.6.1. Central-western accentual type

Present in most of Guipuscoa along with some areas in southeastern Biscay and western Navarre, stress on the second syllable is the most common type of stress. It has also been proposed by Michelena to be common to all varieties in older stages of the language (1988 [1957/58], *FHV*: 405-410). In most modern varieties with this type of accentuation, disyllabic words are stressed on the initial syllable instead, thus avoiding final prominence: *emákumea* 'the woman', *gizóna* 'the man', *mutíla* 'the boy' but *néska* 'the girl'. There are also a number of marked words with initial stress, mostly borrowings and compounds (Hualde *GB*: 71): *básoa* 'the drinking glass' (< Sp. *vaso*, cf. *basóa* 'forest'), *léngusua* 'cousin', *béstea* 'the other', *égia* 'the truth', etc. In many varieties, plural forms also bear initial stress whenever they are either disyllabic or trisyllabic, but maintain post-initial accent when longer: *gízonak* 'the men', *gízonai* 'to the men', but *emákumeak* 'women', *emákumei* 'to the women'.

4.6.2. Northern Biscayan accentual type

In this region there is a lexical distinction between unaccented and accented words. This distinction is morphological, in the sense that roots and affixes are lexically specified as accented or unaccented. Most native roots and almost all singular affixes are unaccented, whereas most borrowings, compounds and plural affixes (as well as a few native roots and affixes) are accented. Unaccented words contain unaccented morphemes, and accented words contain one or more accented morphemes. Unaccented words do not bear prosodic prominence on any syllable, except when they occur in isolation or immediately preceding the verb, contexts in which they surface with final prominence in most local varieties. Accented words always display stress on a given syllable. In most subvarieties of Northern Biscayan Basque, an accent is assigned to the syllable preceding the accented morpheme,[56] and in a few subvarieties stress is assigned to the penultimate or antepenultimate syllable.

Examples of unaccented words are *semié* 'son', *mutillerí* 'to the boy', *lagunentzakó* 'for the friend', and examples of accented words are *béste* 'other', *léku* 'place', *léngusu* 'cousin', *egúzki* (or *éguzki*) 'sun', morphologically derived words such as *edérrago* 'more beautiful', *báltzegi* 'too black', *lóti* 'sleepy', or plural forms like *sagárratatik* 'from the apples', *itturríetara* 'to the fountains'.

56 When there is more than one accented morpheme, it is the syllable preceding the leftmost morpheme that bears stress.

Accent in Northern Biscayan Basque is realized as a H*+L falling contour tone on the accented syllable (Hualde 1989, Hualde & Elordieta, G. & Elordieta, A. 1994, Elordieta 1997, 1998, Gussenhoven 2004: 171, among others).

Recently, the High Navarrese variety of Goizueta has been found to be of the pitch-accent type, with similarities to Northern Biscayan Basque (cf. Hualde & Lujanbio 2008; Hualde & Lujanbio & Torreira 2008), showing that the prosodic system observed in Northern Biscayan Basque must have extended over all the western areas in older times (cf. Hualde 1993a, 1995, 2003a, 2006, 2007, Elordieta 2011a, 2011b).

4.6.3. Souletin accentual type

The Souletin dialect has a different accentual pattern, which is not far from that of the extinct Roncalese dialect. A similar pattern may also be hypothesized for older stages of Salazarese, although not for the modern dialect (*FHV*: 397-398). These dialects have their unmarked stress on the second syllable from the right (i.e., most words are paroxytones) as in Souletin *gízun* 'man', *gizúna* 'the man', etc., and there may have been a time when every word was stressed on the penultimate syllable.

Nevertheless, this pattern is not entirely systematic today. In those words in which a final vocalic coalescence has occurred, both modern Souletin and Roncalese have word-final stress: *alhába* 'daughter' + abs. sg. *-a* > *alhabá* 'the daughter', *néska* + *-a* > *neská* 'the girl'. This also occurs in non-absolutive plural forms (i.e., those bearing an *-e-*) such as S. *gizunék* 'the men (erg.)', *gizunén* 'of the men', *gizunér* 'to the men' (< **gizon-a(g)-ek*, **gizon-a(g)-en*, apud *FHV*: 394, etc.). Alongside these exceptions, oxytonic stress is found on some derivational suffixes (*izabár* 'from Izaba', *larrañtár* 'from Larraine', *aitañí* 'grandfather' < *áita* 'father'), compounds with a monosyllabic second member (*erháuts* 'dust', *ollár* 'rooster') and some other words, both in the borrowed vocabulary (*biarnés* 'Bearnese', *repattán* (cf. Pyrenean Aragonese *repatán*) 'young shepherd', *arrés* 'sheep', *animál* 'huge') and native vocabulary, mostly ending in diphthongs (*añái* 'ram' (< older *añári*), *exéi* 'vixen' (< *azeri*), *aihái* 'dinner' (< *aihari*), S. *ardű*, R. *ardáũ* 'wine' (< **ardano*), *orgá* 'cart' (< **organa*), *etsái* 'foe', *eztéi* 'wedding', R. *idói* 'swamp', *odéi* 'cloud', etc.)[57] (*FHV*: 394-397,

57 Michelena (*FHV*: 397) notes the large number of stressed final diphthongs in Souletin and proposes that they originated from the loss of an intervening consonant, which he hypothesizes to be an aspiration, given its restriction to the first two syllables in the modern dialects.

Hualde *GB*: 72). Such words, at least those in the inherited lexicon, may also be the consequence of resyllabification after the disappearance of the final syllable.

Two main factors distinguish the Roncalese accentuation from that of Souletin (*FHV*: 395). Firstly, in Roncalese, declinational suffixes do not affect the location of the stress, as long as they do not imply any syllable loss: *gízon* 'man', *gízona* 'the man', *gízonaren* 'of the man', *gízonareki* 'with the man', but cf. *alába* 'daughter' vs. *alabá* 'the daughter'. Secondly, Roncalese adds some proparoxytones to the oxytones and paroxytones present in Souletin. This accentual retraction only seems to affect paroxytones —since words that have developed oxytonic stress tend to remain so— and applies not only to native words but also to paroxytonic loanwords. Examples from Izaba are *bézino* (cf. Sp. *vecino*) 'neighbor', *kósino* 'cousin' (*kosíno* in Uztarroze), *ézkapa* 'escape, run away', *zámari* (< Lat. *sagmārius*) 'horse', etc. Michelena (*FHV*: 397) relates these accentual differences to the lack of syncope, which is usual in Roncalese (cf. §4.3.4.5), and to the resyllabification of mid and high vowels (§4.3.4.2) in Roncalese and Salazarese, where *-ea* and *-oa* are monosyllabic.

4.6.4. Old Accentuation

As we have mentioned earlier in this work (in §4.4.3.1.1 and §4.4.2.3), Michelena (*FHV*: 407) proposed that stress was located on the second syllable of the word (as opposed to Martinet 1974 [1950], who proposed that it was on the first syllable) since this would account for the modern distribution of both the aspiration and the aspirated stops in the dialects that still maintain them. Aspiration is limited to the first two syllables in a word, and Michelena relates this to the presence of stress in the second syllable, so that the generalization would be that aspiration occurs on the stressed syllable or the one preceding it.

However, as Michelena himself stated (*FHV*: 418), assuming peninitial stress would run into the problem posed by the vowel neutralizations and apocopes found in the second syllables of disyllabic first members of compounds, in which final *-a*, *-e* and *-o* are neutralized to *-a* —probably after having been produced as a central unstressed schwa— and *-i* and *-u* are dropped (cf. also longer words, in which all final vowels are lost, see *FHV*: 410). In order to prevent this incompatibility, Michelena hypothesizes that peninitial stress was more modern than the system with initial stress proposed by Martinet. This would avoid any conflict with the neutralization or loss of a vowel that would have been stressed if the demarcative stress on the second syllable were already operative. The proposal of a late date for this stress pattern may also account for the aspirations found in the documents from San Millán originally written in the 11th century,

which appear on any syllable of the word. Lakarra (2005, 2006b) and Igartua (2001) also consider the period with initial stress to be posterior to the period with second-syllable stress.

Elordieta (2011a, 2011b) proposes two possible diachronic evolutions of Basque stress and explains the transition from one stage to the next. Elordieta (2011a, 2011b) begins with a mainly monosyllabic stage in which the first disyllabic words were created by means of reduplication. At that time, stress would have fallen on the last syllable, which was also the second. Elordieta subsequently provides support from tonal and intonational phonology to account for a later stage in which stress was word-initial (as proposed by Martinet 1974 [1950] and also assumed by Michelena in *FHV*) and for a further evolution into phrase-final stress, as in Northern Biscayan Basque. However, following an argument by Martínez-Areta (2004), Elordieta points out a problem for assuming word-initial stress, and that is the impossibility of explaining why marked words in many central varieties are characterized precisely by having initial stress (cf. Elordieta 2011a: 1006-1007). That is, if all Basque words were initially stressed at some point, it would be impossible to distinguish marked words on the basis of initial stress. Thus, Elordieta proposes a chronology without the conflicting stage, with the older word-final prominence becoming penultimate stress in eastern varieties and phrase-final in western dialects. Only Northern Biscayan Basque maintains the phrase-final stress system, and all other western stress systems derive from the old phrase-final stress system.

Hualde (2007) proposes that the accentuation pattern of Proto-Basque could have been similar to the pattern present in modern northern Biscayan, and that the pitch rise on the second syllable of the word may have been reinterpreted in eastern varieties as a demarcative non-contrastive stress on the second syllable. Hualde (2003a) claims that the influence of Spanish stress should be considered when understanding this reinterpretation, since this is also realized as a pitch rise. This hypothesis was confirmed empirically by Elordieta & Hualde (2003) by means of a perception test. The historical relation between the Northern Biscayan and the central-western accentual type is also favored by the fact that they share the difference between stressed and unstressed —or marked and unmarked— words, and the aforementioned divergence in the realization of accentuation (pitch accent vs. stress) is a typologically frequent change.

According to Michelena (*FHV*: 423-424), the transition from the central accentual system —reconstructed by him as the common accentuation— to the modern Souletin paroxytonic stress would have been the consequence of a reinterpretation of the position of the stress in trisyllabic words as counted from the right (instead of the left) edge of the word. This would be followed by a subse-

quent regularization of the new pattern in disyllabic words, which would have their stress systematically relocated to the initial syllable, i.e. the penultimate.

Let us roughly follow Hualde in order to accept the aforementioned chronological sequence, and propose that the diachronic evolution must have been phrase-final stress (nowadays only found in Northern Biscayan Basque) > peninitial stress (in central varieties) > penultimate stress (in eastern varieties). In this way, we can discard the reconstructed old second syllable stress proposed by Michelena and accepted by most authors. We must assume that the loss of the aspiration after the second syllable occurred not generally but only eastwards from the central territory, in an area which includes precisely those dialects in which the /h/ is maintained to this day. This would not be incompatible with the appearance of the aspiration on any syllable in medieval documents in the meridional area, given that such varieties probably followed the western accentual model. However, a reconstructed prefixing, mainly a monosyllabic stage with stress on the second and final syllable —occupied by the root— seems rather plausible (cf. Lakarra 1995, 2006b).

Regarding the proposal of demarcative stress on the first syllable (Martinet 1974 [1950]), I agree with Elordieta (2011a) that it is probably the hypothesis with the least empirical support.[58] Martínez-Areta (2004) rejects this hypothetical phase of the accentuation altogether. I can point to another weakness in Martinet's proposal, namely that the prosodic prominence granted to this position (in order to explain the consonantal processes affecting stops that occurred word-initially) can be attributed to word-initial position (cf. Smith 2005), without having to posit the presence of stress on that syllable.

4.7. Conclusion

Since the seminal work of Michelena (*FHV*) on the historical phonology of Basque, some aspects of the diachronic evolution of Basque phonology have received considerable attention by different researchers in the last two decades. These include the inventory of stop consonants in Old Basque (cf. §4.2.2) and the reconstruction of the accentuation system (cf. §4.6.4). Here we have provid-

58 The most frequently-mentioned fact which favors initial stress is that it would easily account for the vowel neutralizations and losses which occurred in the second and third syllables of the first members of compounds mentioned supra. However, Hualde (2006, 2007) gives a segmental explanation for such neutralizations: according to his hypothesis, the addition of an intervening laryngeal segment in composition would not only explain the vowel reductions but also the devoicing and neutralization of the stops, which become alveolar /t/, perceptually the most similar segment.

ed a revision of Michelena's work in the light of these recent advances, in an exercise that has tried to put this author's influential proposals into perspective. Revisiting Michelena's work while taking advantage of the progresses made by both typological approaches to phonology and phonetically-based phonology has proven fruitful. Without trying to make any significative incursion into the field of phonetic explanations for phonological processes, a couple of proposals have already surfaced in the development of this work, e.g. a hypothetical relation for the allegedly unconnected contexts in which the otherwise general /u/ > /y/ fronting fails to develop (cf. §4.3.3) and the reasons behind the metathesis of sounds bearing elongated phonetic cues (as understood by Ohala 1993), which does not apply to segments lacking such features (cf. §4.4.5.3). This can only encourage us to advance in the task of applying this type of framework to the Basque diachronical data.

5. Root Structure and the Reconstruction of Proto-Basque

Joseba A. Lakarra
(University of the Basque Country, UPV/EHU-JUMI)

5.1. Introduction*

In Lakarra (1995) and subsequent works, I have shown that the reconstruction of Proto-Basque (= PB) carried out by Michelena in his essential works of the 1950s can be expanded and deepened. Thus, through the theory of the monosyllabic root, we can extend our reconstruction to stages prior to Michelena's PB, currently considered standard by the great majority of Bascologists.

The main task of the following synthesis lies precisely in the revindication of the fundamental role which the theory of the root —including the analysis of the structure and the evolution of its canonical form— plays in both the reconstruction and the comparison of any language. The stress on this point has provided a methodology which has proven extremely useful for the analysis of the numerous fundamental features of the system (as well as its changes over time) of a number of languages and language families with a long tradition of research and undeniable results, including Indo-European, Semitic, Uralic, Austronesian, and Sinitic. I believe that this methodology is also applicable to the diachronic study of Basque and PB. In addition, some paths which have remained unex-

---

* "Historia de la Lengua Vasca y Lingüística Histórico-Comparada" (HLMV-LHC, Ref. GIC 10/83, IT 486-10), "Monumenta Linguae Vasconum (II-III-IV)" [= HUM 2005-0847, FFI2008-04516, FFI2012-37696], and "Hizkuntzalaritzalaritza Teorikoa eta Diakronikoa: Gramatika Unibertsala, Hizkuntza Indoeuroparrak eta Euskara" (UFI 11/14).

This is an abbreviated and revised version of Lakarra (2011a), on which the paper of the same title read at the *2nd Congress of the Michelena Chair* (Lakarra 2011e) was based. In Lakarra (2013), I present new paths —concerning both chronology and grammaticalization— to broaden and deepen our knowledge of Proto-Basque. I want to thank the members of the audience at this presentation, and extend these thanks to Borja Ariztimuño, Ricardo Gómez, and Ander Egurtzegi for their careful readings of an earlier version of this work, and to Mikel Martínez-Areta for translating and editing the text.

Abbreviations: C = consonant, V = vowel, T = stop, R = sonorant, S = sibilant, PB = Proto-Basque, CB = Common Basque. By PB, I mean the earliest period that is reconstructible to some degree. Although its chronological ubication can only be determined in structural terms (before CB), it might be reasonable to date it to the last few centuries BC. As for CB, it is the last stage of the language prior to its dialectal fragmentation. Based on a variety of reasons, Michelena (1987 [1981a]) dated this CB to around the 5th-6th cc.

plored in Michelena's reconstructing paradigm, as well as others provided by the typological research of recent decades, offer us the possibility to go, through the theory of the monosyllabic root, even further back in the history of Basque.

This chapter will be organized as follows: §5.2. On the standard reconstruction of PB; §5.3. Reasons for change; §5.4. A new theory for the reconstruction of Proto-Basque: the theory of the monosyllabic root; §5.5. Theory of the root and formal etymology in the Basque lexicon; §5.6. From the canonical form to diachronic holistic typology; §5.7. Canonical form and reconstruction of the Basque verb; §5.8. Canonical form, word families, and reconstruction of the Old PB phonology; §5.9. Changes in the canonical form; §5.10. Some conclusions and unresolved issues; §5.11. Appendices (§5.11.1. Analysis of the CVC root pattern and the bisyllabic root patterns; §5.11.2. Some representative examples of new etymologies).

5.2. The standard reconstruction of Proto-Basque

Numerous attempts have been made to relate Basque to a considerable number of different languages or, more precisely, to establish a genetic relationship between them (cf. Michelena 1988 [1964b], Trask *HB*, last chapter). Despite this, progress in the research on the prehistoric evolution of the Basque language and the structure of Proto-Basque has only been made thanks to an attempt at an internal reconstruction of the old stops (Martinet 1974 [1950]) and, above all, to the immense work carried out by Michelena, crowned by his magnificent *Fonética histórica vasca* (1977 [1961]), along with certain other works including "Las antiguas consonantes vascas" (1988 [1957a]) and *Sobre el pasado de la lengua vasca* (1988 [1964b]).

According to Michelena's reconstruction (cf. §4.2.2), the stop system postulated for Proto-Basque —defined as the state of the language at the time of the arrival of the Romans, in the 2nd century BC— would resemble the one shown in Table 5.1.

Table 5.1. Consonant system of Proto-Basque according to Michelena

Fortes	__	T	K	L	N	R	Ś	S
Lenes	b	d	g	l	n	r	ś	s

The phonological system would be completed with the addition of: /h/, five vowels (*a*, *e*, *i*, *o*, *u*), and five descending diphthongs (*ai*, *au*, *ei*, *eu*, *oi*). Fortis stops were pronounced with a marked aspiration in word-initial position, and

were more tense and longer than lenis stops in word-medial position. Lenis stops had no aspiration, and would be phonetically spirant in word-medial position. All stops were phonologically voiceless. Sibilants and sonorants would be neutralized in favor of the lenes in word-initial position, and in favor of the fortes in word-final position.[1] Cf. §4.2.2, for more details.

In general, vowels have been very stable in terms of timbre, except for some phenomena related to word derivation and word compounding. Here, between the Aquitanian period (1st-3rd centuries AD) and the first medieval documents (9th-10th centuries), *-a*, *-e*, *-o* are neutralized into *-a* in the last syllable of the first member of two syllables, and *-i*, *-u* drop out in the same context. In the last syllable of a first member of three syllables, any -V drops out. By contrast, diphthongs are very unstable, especially those ending in *-i*, and they show a tendency to increase in number over history. Besides, [wi] can only really be considered to be a (lexically not very widespread) variant of /oi̯/.[2]

The syllabic structure of Old Basque was defined by Michelena (1985 [1979a]) as (C)V(W)(R)(S)(T) and reduced to CVC by Artiagoitia (1990). As we can see, only vowels can be a syllabic nucleus,[3] and consonant clusters are not accepted in word-initial position. In the patrimonial lexicon, rhotics and affricates cannot appear in word-initial position either. As for the syllable-final position, consonant clusters can only appear in the following order: sonorant-sibilant-stop, actually corresponding to the sonority hierarchy (cf. Iverson & Salmons 1992). As Michelena observed, no word is known which contains a syllable in which all slots are occupied, and Artiagoitia noticed that -T (= stop) always corresponds to late variants in noun compounds (*art-* from *ardi* 'sheep' in *artalde* 'sheep flock'), or to the devoicing of a -D#, arisen after the apocope of the vowel of an original *-DV# (kept in non-word-final allomorphs): **du-da* → *du-da-la* 'that (I) have' ~ **du-da* > *du-t* 'I have'.

1 Given the secondary importance that has been assigned to the issue of aspiration (cf. Trask *HB*; against this view, cf. Igartua 2001, 2006, 2008, 2011, and Lakarra 2009c, 2009d, 2010b, 2011c, 2011d), it is important to emphasize that the PB reconstructed by Michelena did have the phoneme /h/. In addition, fortis stops were allophonically pronounced as aspirated in word-initial position, and so were the aspirated voiceless stops in word-medial position at a later stage in the language. Recent research has drastically reduced the number of non-etymological /h/, and shown that, for Michelena, the cases yet to be explained of dropping of /h/ were many more than the cases of non-etymological aspirations (cf. towards the end of §5.4, in footnote).

2 For the reconstruction of such a system, both the analysis of Latin and Romance loan words and the structural analysis carried out by Martinet and Michelena were essential (cf., on the latter, Michelena 1988 [1957b], 1987 [1974]).

3 Cf., for a very different situation, what happens in Berber and Kartvelian languages, once alleged to be possible relatives of the Basque language (cf. the end of §5.4).

Martinet himself pointed out that it was the observation of evolutions such as the following ones which made him rule out a voiceless/voiced opposition for stops, like the one present in Latin and Romance: Lat. *pice* > Basq. *bike* 'pitch', as well as Lat. *ballēna* > Basq. *balea* 'whale' (i.e. initial voiceless stop > voiced stop, vs. voiceless > voiceless in Romance), but Lat. *sēta* > Basq. *zeta* 'silk' (i.e. intervocalic voiceless stop > voiceless, vs. voiced in Romance). He then resorted to the fortis/lenis opposition, similar to the one that he had already postulated for Danish.

Michelena (1988 [1951b]) strengthened the case for this argument by pointing to the phonetic behavior of [f], late and an allophone of both /b/ and /p/, contrary to what would be expected in a language with a voiceless/voiced opposition as its primary feature.[4] The assumption that the primary feature of the opposition was fortis/lenis was also borne out by the fact that geminate voiced stops and other Latin and Romance voiced stop clusters are borrowed as voiceless stops in Basque, but develop as voiced in Romance: Lat. *sabbatu(m)* > Basq. (B) *zapatu* (~ Sp. *sábado*), Rom. *cobdicia* > Basq. *gutizia* (~ Sp. *codicia*). According to Michelena, this occurred because the relevant feature in Basque was fortis/lenis at an old period. Thus, to Basque ears it was irrelevant as to whether a geminate such as *-bb-* should be voiced or voiceless. The relevant feature was that it was long, hence also stronger, i.e. fortis, and as such it was reproduced.[5] A third relevant argument is the abundance of graphic geminate voiceless stops (interpreted as non-aspirated fortes by Michelena) in Aquitanian, in contrast to the inexistence of geminate voiced (lenis) stops.

Since sonorants and sibilants showed the same neutralization positions within the word, the fortis/lenis analysis was later extended to them (cf. Michelena 1988 [1957a]). Phonological contrasts occurred in word-medial position alone, with lenes (= fricatives) occurring in initial, and fortes (= affricates) in final position. It is interesting to observe that Latin geminate laterals and nasals are simplified when they enter the Basque phonological system, whereas the simple Latin ones undergo lenition: Lat. *angelu* > Basq. *a(i)ngeru* 'angel', Lat. *angellu*

4 On /f/, cf. Hualde (1997a), where **wh* > *f* is postulated, Lakarra (2009c), and the Appendix §5.11.2, which contains several etymological proposals (*afari*, *ifar*, *ifini*) based on this change.

5 For possible alternatives —exclusively limited to stops—, see Trask (1985) and Hualde (1997a), and, for a criticism of both, Lakarra (2008c). Let it suffice to say that both put forward data which are hardly PB, but 15 or 20 centuries later. Furthermore, they do not present any etymology which could justify their new reconstructing proposals. In any case, Trask (*HB*) explicitly states that he has given up his previous view, and returns to the Michelenian system, even though this does not seem to include the treatment of aspiration (cf. the bibliography cited some footnotes further up).

> Basq. *angelu* 'angle', Lat. *annōna* > Basq. *anhoa*[6] 'daily ration', Lat. *arēna* > Basq. *harea* 'sand'. This is also the case with words of patrimonial origin: **ardano* > *ardo* 'wine' (~ dial. *ardãõ*), etc., but **arraNo* > *arrano* 'eagle'.

As for sibilants, it is worth noting that, in the oldest loan words, Latin /s/ enters Basque as a dorsal: Lat. *corpus* > Basq. *gorputz* 'body', Lat. *sēcula* > Basq. (S) *zekürü* 'time, lifetime' (the general *sekula* is a late loan word, as *-l-* also shows), *baradizu* 'paradise' / *paraisu* (both are Basque variants; note also *b-*/*p-*), etc. In the realm of rhotics, they are not allowed in word-initial position, and hence in loan words they are avoided by means of a prothetic vowel (Sp. *ropa* > Basq. *arropa* 'clothes', Sp. *ratón* > Basq. *arratoi* 'mouse', etc.). In word-final position, neutralization is historically almost complete, apart from a few cases such as *hur* 'water' / *huR* 'hazelnut' (in Roncalese), *plateR* 'dish' ~ *platera* 'the dish', etc.

5.3. Reasons for change

Before 1995, Bascologists showed very little interest in studying the canonical form of the root and morphemes. Thus, in his *Morfología vasca* (1969 [1923-25]), of over one thousand pages, Azkue devotes barely 10 lines to the root (which, incidentally, he considered to have been monosyllabic in the past).[7] Lafon's short study (1999 [1950]) is only a review of Uhlenbeck (1947 [1942]), along with a revindication of the alleged role of the Basque-Caucasian comparison, and an explicit confession of an almost complete mistrust of Basque internal reconstruction.[8]

---

6 Cf. Lakarra (2011c, d) and *unhai(n)* 'oxherd' in §5.11.2, on the *-h-* of *anhoa* 'daily ration' (which initially came from the /h/ arisen in the 3rd syllable (= h_3) as a result of the lenis /n/), as well as that of *elhe* 'word, fable', *unhatu* 'tired', *olho* 'oats', etc.

7 An even earlier forerunner of the same idea is Aizkibel (in the 19th century), rescued by Gómez (2008a).

8 Admittedly, this can be more justifiable at that time than from 1961 onwards, when Michelena's *FHV* had been published. Here are his words:
"Un nom basque, s'il ne contient aucun préfixe ou suffixe connu et dont la valeur soit nette, ne se laisse généralement pas analyser; ou, plus exactement, on ne possède aucun moyen direct de l'analyser. Mais on peut essayer de recourir à un moyen indirect, qui consiste dans l'emploi de la méthode comparative. Il faut chercher des mots appartenant à la même racine dans des langues apparentées génétiquement à la langue basque, c'est-à-dire dans les langues caucasiques" (Lafon 1999 [1950]: 508).
Let it be recalled, in any case, that he himself (1951-1952) admitted the weakness of such a comparison, and was aware that Basque-Caucasian reconstructions had been unable to

Michelena's reconstruction brought about a huge advance in diachronic research, but it was purely segmental. That is, a phoneme was reconstructed in a proto-form by starting from the triangulation of the phonemes found in historical cognates (cf. Durie & Ross (eds.) 1996, Ringe 2003).[9] There is hardly any reference to the structure of the root or the morpheme in the *FHV*, even though taking it into account could have led him to satisfactory explanations for processes such as the assimilations or dissimilations of homorganic sibilants (Basq. *sasoi-z* 'in time' < Sp. *sazón*, Basq. *solas-ez* 'talking' < Sp. *solaz*, Basq. *sinetsi-z* 'believing' ~ Dch. *zin-hetsi-z* (apud *FHV*), etc.) and rhotics (Basq. *elur/edur* 'snow' < *erur*, Basq. *belar/bedar* 'grass' < *berar*, etc.), which clearly occur in this context.[10]

In fact, Michelena was conscious of the relevance of approaches such as Benveniste's (1935) theory of the root in the reconstruction of Indo-European, as we can read in *Lenguas y protolenguas* (1990 [1963]), but he did not aim to apply a similar procedure to his own work. In these, we find monosyllabic, bisyllabic and polysyllabic roots, although he stated more than once that it was the group of bisyllabic roots which corresponded to the oldest stratum of the language.

Along the same lines, Uhlenbeck (1947 [1942]) had established an extensive relation of old radical patterns,[11] with the purpose of isolating the different strata (Caucasian, Mediterranean, Pyrenean, etc.) which, according to him, were present in the language. He identified the following patterns:

a) monosyllables:
    1) vowel-ending (*khe* 'smoke', *su* 'fire');
    2) diphthong-ending (*gai* 'thing, able', *gau* 'night');
    3) the "much more numerous" consonant-ending ones (*har* 'worm', *gar* 'flame', *hil* 'to die, to kill, dead', *lan* 'work');

---

make any progress worth mentioning. Michelena (1988 [1964b], last chapter) and (1985 [1968]) were the *coup de grâce* to this approach.

9 Michelena proposed around 1400-1500 etymologies (cf. Arbelaiz 1978). His interest in morphological reconstruction was late and clearly secondary, but in addition to multiple observations made in the *FHV*, *Apellidos Vascos* (Michelena 1997 [1953a]), and some other works, Michelena (1987 [1971b]) and (1987 [1977]) are essential contributions to this field.

10 He did not forget, however, the accent (cf. Michelena 1988 [1957/58], 1987 [1972c], as well as scattered notes in many other works, and the chapters on accent in the *FHV*, somewhat expanded in the second edition). Cf. now also Martínez-Areta (2004) and Elordieta (2011a). I also referred to the need to postulate an accent on the 2nd syllable (i.e. on the monosyllabic root; cf. Donegan & Stampe 1983, 2004) for Old PB periods in Lakarra (1996b) and (2006b), which would be coherent with a [pref.-root] structure.

11 Cf. §5.7 for some alleged verbal roots proposed by Lafon (1980 [1944]).

4) the numerous affricate-ending ones (*haitz* 'rock', *antz* 'resemblance', *hats* 'breath', *gaitz* 'big, misfortune');
5) reduplicate forms, not very numerous (*konkor* 'hump', *gogor* 'hard', *garagar* 'rye');

b) bisyllables:
1) VCV and CVCV, numerous and much more characteristic of Basque (Uhlenbeck 1947 [1942]: 559): *habe* 'column', *aho* 'mouth', *ale* 'copy, unit', *harri* 'stone', *begi* 'eye', *behi* 'cow', *beso* 'arm', *xori* 'bird', *eme* 'female', etc.;
2) VCVC and CVCVC, "very common": *abar* 'small stick', *adin* 'age', *behor* 'mare', *behar* 'duty', *gizon* 'man', *sagar* 'apple', etc.;
3) VCCV and CVCCV, "very common": *albo* 'side', *argi* 'light, clear, clearly', *garbi* 'clean', *zaldi* 'horse', etc.;
4) VCCVC and CVCCVC, "also very common": *argal* 'thin', *azkar* 'strong, intelligent', *bazter* 'corner', etc.;
5) with diphthong (DCV, CDCV, DCVC, CDCVC, VCD, CVCD, VCCD, CVCCD, VCDC, CVCDC, VCCDC, CVCCDC): *aita* 'father', *leiho* 'window', *oihu* 'cry', *aihen* 'shoot', *leizar* 'ash tree', *etsai* 'enemy', *garai* 'period, high', etc.

c) trisyllables: Uhlenbeck considered the types VCVV or CVCVV to be rare: *egia* 'truth', *herio* 'death (active sense)', *lehia* 'competition'.[12]

As is well known, reconstruction based on the analysis of the root has been of the utmost importance in the Indo-European and Semitic families. Benveniste (1935) and some other authors detected regularities and restrictions in the phonotactic combinations of morphemes.[13] These can only be explained away by

12 Uhlenbeck stated that he was not interested in individual etymologies, but in root patterns. See §5.5, and the Appendix in §5.11.1 (% of attested roots (Table L), % of fossils which can be candidates to PB. roots (Table H) and % of loan words (Table M)).

13 The first passage of Benveniste's PhD is quite revealing in this respect:
"L'objet essentiel de la grammaire comparée, depuis une soixantaine d'années, a été de poser des correspondances entre les langues indo-européennes, et d'expliquer, en parlant de l'état qui définissent ces correspondances, le développement des dialectes attestés. On met donc au compte de l'indo-européen tout ce qui semble hérité dans chacun des dialectes, avec la conviction, implicite ou avouée, qu'on ne saurait sans danger pousser la restitution au delà du prototype immédiat. Depuis le *Mémoire* de F. de Saussure, le problème de la structure des formes indo-européennes elles-mêmes a été presque complètement négligé. Il paraît communément reçu qu'on peut analyser l'évolution de l'indo-européen sans se soucier de ses origines, qu'on peut comprendre des résultats sans pousser jusqu'aux principes. De fait, on ne va guère au delà de la constatation. L'effort, considérable et méritoire, qui a été employé à la description des formes n'a été suivi d'aucune tentative sérieuse pour les interpréter. Là est sans doute la cause principale du

starting from a previous phonology and grammar, whether closer or further away from the historical period. More recently, something similar has been found for Uralic (Bakrò-Nagy 1992), Austronesian (Blust 1988), Chinese (Sagart 1999), Tibeto-Burman (Matisoff 2003), and certain other languages and language families.[14]

> "Ce qu'on a enseigné jusqu'ici de la nature et des modalités de la racine est au vrai, un assemblage hétéroclite de notions empiriques, de recettes provisoires, de formes archaïques et récentes, le tout d'une irrégularité et d'une complication qui défient l'ordonnance. On enregistre des racines monosyllabiques (**bher*-) ou dissyllabiques (**g*^{w}*eyə*); des racines bilitères (**dō*-), trilitères (**per*-), quadrilitères (**leuk*-), quinquilitères (**sneig*^{w}*h*); des racines à voyelle initiale (**ar*-) ou à voyelle finale (**pō*-); à degré long (**sēd*-) ou à degré zéro (**dhək*); à diphtongue longue (**srēig*-) ou brève (**bheudh*-), à suffixe ou à élargissement, etc. On serait en peine de justifier et même d'énumérer complètement tous les types de racines qui sont attribués à l'indoeuropéen. Il y a ici un abus de mots qui trahit une doctrine indécise. On n'obtient pas de l'indo-européen en additionnant les diverses formes indo-européennes d'un thème verbal ni en projetant dans la préhistoire des particularités d'un état de langue historique. Il faut essayer, par de larges comparaisons, de retrouver le système initial sous sa forme la plus simple, puis de voir quels principes en modifient l'économie" (Benveniste 1935: 147-148).

Although Michelena considered Benveniste his master, and deeply admired his reconstruction based on the canonical form of Indo-European roots (cf. Michelena 1990 [1963]), he did not apply this analysis to the Basque language. If we consider a few examples of his reconstructing work, e.g. through the compilation made by Arbelaiz (1978) of his etymologies, it is clear that Benveniste's criticism of the Indo-European research prior to his own research, is likewise applicable to Michelena's work itself, point by point and without major changes. Among his reconstructions, we find bisyllables, monosyllables, and polysyllables, vowel-initial and consonant-initial words as well as vowel-ending and consonant-ending words, various consonant groups in both word-medial and word-

---

malaise actuel de la grammaire comparée: si la recherche proprement comparative tend à s'éparpiller en travaux de plus en plus menus, c'est qu'elle a oublié les questions fondamentales; et si bien des linguistes se détournent de la comparaison, c'est pour s'être laissés aller à croire que l'on n'avait plus de choix qu'entre le connu et l'inconnaissable" (Benveniste 1935: 1).

14 In Semitic, the canonical form of the root (and its implications, such as the certainty as to the common origin of the root structure of all individual languages) is so evident —and common to languages such as Arabian and Hebrew— that according to a number of scholars, it is this obviousness itself which has delayed the comparison between the different languages of the family, particularly among the aforementioned ones.

final position, etc. All this betrays the absence of any theoretical pattern, and even of any concern as to the canonical form of PB. roots.

Michelena (1990 [1963]) considered the canonical form of morphemes to be typical of the deepest reconstructions, but he did not manage to use —or to specify— this canonical form when elaborating his own etymologies. As a result, in these we find a great diversity of proto-morphemes, allomorphs and patterns, perhaps also corresponding to different periods. They are a kind of compilation of reconstructions (by the addition of proto-phonemes) of morphemes which are historically attested. Hence, they should not be the point of arrival of any systematic reconstruction, but rather its point of departure. The new reconstruction is based, not on the analysis of loan words —which can only cast light upon more recent periods (cf. Lakarra 2012c)—, but on the irregularities of patrimonial words, which belong to an earlier period. If we adopt a more restrictive syllabic structure than that of Michelena, and consequently point to hitherto unnoticed limitations such as **TVTV —or even **CVCV and **bisyllabic—, we must necessarily postulate a root pattern (monosyllabic and of a CVC structure) which accounts for such restrictions, ignored by the earlier paradigm, and unexplained —if not unaccountable— within it.

5.4. A new theory for the reconstruction of Proto-Basque. The monosyllabic root

In Lakarra (1995), I revised some restrictions on old syllables and roots, and pointed out the absence of the root structure TVTV. We would expect this structure to be more frequent in a theoretically bisyllabic language, since stops are historically the most abundant consonant type in the language. In fact, there exist several dozens of words with that structure. The total number, however, is far below the potential 900 (6 x 5 x 6 x 5). And, more significantly, all of those words can be analyzed as one of the following: a) loan words (*bake* 'peace', *baba* 'broad bean'); b) compound or derived words (*bada* 'thus, then', *bage* 'without', *begi* 'eye'); c) onomatopoeias, phonosymbolisms and nursery words (*babo*, *baka*, *bebe*, *bobo*, *buba*, etc.). To this, it should be added that, leaving aside phonosymbolisms and grammatical elements (*ni* 'I', *hi* 'thou', *gu* 'we', *zu* 'you (sg., actually < pl.)', *ote* 'interrogative particle' (a Latin loan), *al* 'interrogative particle', *edo* 'or', *ala* 'or (disjunctive)'), almost all of the historically attested monosyllabic lexemes have a CVC structure: *hats* 'breath', *zur* 'wood', *gor*

'deaf', *lur* 'earth', *bil* 'to gather', *bits* 'foam', etc. Lexical roots without any onset (**VC) or coda (**CV) are not allowed.[15]

I therefore proposed, in 1995, that the old lexical roots of Proto-Basque were monosyllabic, and of a CVC structure. This proposal could account for all of the above mentioned restrictions (**CV, **VC, **TVTV). However, in turn, the need arose to provide an explanation —through derivation,[16] compounding, or borrowing— for the great majority of the Basque historical lexicon, consisting of multiple bisyllabic and (to a lesser extent) trisyllabic and tetrasyllabic patterns.

In that work of 1995, and in some later studies such as Lakarra (1998a) and (2008c), several results were drawn at different levels, which clearly supported the new proposal. Here are some of them:

a) New examples of combinations were found, in which previously known roots were present:
   - **bel* (we already knew *harbel* 'slate', *ubel* (< **hur-bel*) 'black and blue', *orbel* (< **hor-bel*) 'fallen leaves', etc.): cf. now *sabel* 'abdomen', *gibel* 'liver; rear part', etc.

b) We identified new roots, hitherto unknown:
   - **ger*: *ok(h)er* 'naughty, crooked', *ak(h)er* 'he-goat', *puzker* 'fart'.
   - **han*: *a(h)untz* 'she-goat', *(h)an-dots* 'ram', *a(h)ari* 'ram'.

c) On formal grounds, loan words were detected, which had been unnoticed until then: *zemai* 'threat', *bazter* 'corner', *beldur* 'fear', *nigar* 'crying', etc.[17]

d) In addition, the new analyses brought new elements to the reconstruction of the old morphology. It definitely had, as did historical Basque, suffixing and compounding. But the new analyzing procedure also allowed us to consider reduplication and prefixing as productive devices of PB, un-

15 Since the phonosymbolic element and the grammatical morphemes are —each in their own way— a kind of mirror image of the lexical canonical form (cf. Gamkrelidze & Ivanov 1984 for Indo-European, Blust 1988 for Austronesian, Bakrò-Nagy 1992 for Uralic, etc.), the fact that both systematically differ from the CVC structure in (Old) Basque constitutes a strong additional argument in favor of CVC as the canonical form of the Old Basque lexicon.

16 At first (cf. §5.8), I considered suffixing to be the only derivational process, but later research has shown that, although PB must have had (as in many other languages) some suffixes, it was prefixing (much more marked cross-linguistically) and reduplication which constituted the most relevant mechanisms for word formation in Old PB.

17 As for the last two, it is evident that **bel*, **dur*, **ni* and **gar* belonged to the PB. lexicon. However, they do not appear to explain anything in *beldur* or *nigar*, and hence we should seriously consider the possibility that these have been borrowed.

known in historical Basque and as yet not proposed in the reconstruction of PB:

a) Reduplication:
   - **dar*: *adar* 'horn, branch' < **da-dar*.
   - **der*: *eder* 'beautiful' < **de-der*.
   - **dol*: *odol* 'blood' < **do-dol*.
   - **nal*: *ahal* 'can' < **na-nal*.
   - **nol*: *ohol* 'plank' < **no-nol*.
   - **zen*: *zezen* 'bull' < *ze-zen*.
   - **gor*: *gogor* 'hard' < *go-gor*.

b) Prefixing:
   - **gi*: *gibel* (cf. supra), *gizen* 'fatty meat', *gihar* 'lean meat', *gizon* 'man', etc.
   - **la*: *labar* 'cliff', *labur* 'short', *lagun* 'mate', *labain* 'slippery', etc.
   - **sa*: *sabel* 'abdomen', *samin* 'bitter, pain', *samur* 'tender', *sabai* 'loft', etc.

On the other hand, the finding of a root structure which was much more precise and restrictive than the previous one made comparisons between Basque and other languages more difficult. This represented a severe blow to the followers of the *bongo-bongo* linguistic method, denounced by Trask (cf. *HB*: last chapter), i.e. the method based on the simple resemblance between words, however superficial and remote this may be, and not on homologies typical of the standard comparative method (cf. Meillet 1925). This method had already been discredited by Michelena (1988 [1964b], 1985 [1968]), who pointed out that the alleged Basque-Caucasian parallels lacked the necessary rigor (such as e.g. regular phonetic correspondences, to be expected whenever we compare two sets of words from languages which are undoubtedly related to each other). Additionally, comparing the respective root structures —the main features of which belong to the *longue durée*— shows us that Basque and Kartvelian stand even further away from each other than Indo-European and Kartvelian could be. In fact, few languages or language families could differ so much from each other in such a significant structural feature (cf. Harris 1990, and Lakarra 1998b).[18]

18 However impossible it may be to demonstrate the absence of a genetic relationship (cf. Hamp 1998) —even through this path of reasoning—, it is worth mentioning that nobody has considered the enormous differences in syllable structure, root structure and related phenomena between Basque and its most famous alleged relatives, such as Bereber and Kartvelian.

5.5. Theory of the root and formal etymology in the Proto-Basque lexicon

Uhlenbeck (1947 [1942]) suggested that it would be more interesting to focus our analysis on the different root patterns, rather than on the etymologies of individual words, due to the regularities which he believed he had detected in some of these patterns:

> "(…) je crois déjà rendre service aux études basque[s] en tâchant de ramener les nombreux radicaux nominaux et les racines verbales de la langue basque restés sans analyse à quelques types déterminés, qui doivent, selon toute probabilité, être considérées comme pyrénéens occidentaux anciens. (…) Pour le moment, il m'est indifférent de savoir si tel élément lexical —mot ou racine— possède ou non une parenté caucasienne plausible. Si pareil élément lexical correspond à un type ancien, il devra passer pour pyrénéen occidental ancien, sauf la possibilité qu'un élément non originellement pyrénéen occidental ancien d'importation ultérieure —mot ou racine— se soit rallié secondairement à un type pyrénéen occidental ancien. Mais cette question est sans importance pour mes investigations, puisque *je recherche non pas des étymologies, mais des types*. Des éléments d'origine étrangère secondairement assimilés à d'anciens types indigènes contribuent, eux aussi à affirmer l'existence de types anciens" (Uhlenbeck 1947 [1942]: 567); the italics are mine [JAL].[19]

Later research has uncovered mistakes of varying significance, in both this and other works by the author. Furthermore, some of his aims, such as his attempt to demonstrate the polygenesis of the Basque language, were clearly on the wrong track.[20] But in spite of all this, and although I do not agree with him on most of his conclusions (particularly on the bisyllabism of the oldest Basque root), I believe that his emphasis on analyzing the Basque lexicon on the basis of root patterns and their evolutions instead of from atomist analyses, should be recognized. Unfortunately, this path of study was not explored until much later (cf. Lakarra 1995 and later works). Thus, for example, in Michelena's crucial

19 Cf. also: "Je me figure les types de mots et de racines que j'ai pu relever dans la langue basque des quatre dernières siècles comme la continuation de types de mots et de racines pyrénéens occidentaux anciens. *Il va de soi que je vise des types et non pas quelque mot ou quelque racine en particulier*. Parmi les mots et les racines que j'ai cités, il y a sans doute un certain nombre d'éléments importés au cours des âges par des peuples étrangers, qui se sont adaptés à des types indigènes" (Uhlenbeck 1947 [1942]: 580) [the italics are mine, JAL].

20 Cf. Michelena (1988 [1964b]) and Lakarra (1986) for the alleged different origin of Biscayan. On the polygenetic ideas of Uhlenbeck, which coincide with those of Boas, Trombetti, Schuchardt, and Tovar, cf. Lakarra (2008c). It is particularly unfortunate that barely ten lines about the different origins of Biscayan and the rest of the dialects have attracted more attention than the rest (over 99 per cent) of Uhlenbeck's ideas on the different root patterns in this important work.

Fonética Histórica Vasca we find only scarce allusions, if any, to the forms of roots and morphemes and their possible relevance for reconstruction.[21]

Since bisyllables constitute the majority of roots in historical Basque, proposing a canonical form CVC for Old PB inevitably leads us to the need to study how the evolution from CVC to the bisyllabic attested patterns (cf. Appendix §5.11.1) could have taken place. In what follows, I sum up the research carried out since Lakarra (1995) on the bisyllabic root patterns, in order to verify whether the decision to exclude these patterns from the earliest PB was sound or not. It should be stated in advance that the current analysis, now far more detailed, could not be more favorable to my hypotheses of over a decade ago.

My first impression is that the attested bisyllabic roots in general, and among them, those which can be considered to be of unknown origin —and hence as candidates for PB—, are very scarce. On the other hand, even though the results for the bisyllabic root patterns are more meagre, and are all very far from the monosyllabic results, it is also possible to establish differences between the bisyllabic patterns, which will undoubtedly correspond to different periods. Among the most recent bisyllabic patterns (CVCV, CVCCV, CVCCVC, CVCVC), we find e.g. any combination of a C plus /f/ or /m/. This type shows extremely few attestations and low percentages, and is made up of onomatopoeias and abundant loan words, but very few fossils. Something similar is found for the pattern **TVTV, which was at the origin of this research. What I merely suspected at first, that this pattern has no chance at all of belonging to PB, can now be asserted based on several types of evidence.

This process requires a high level of formal precision, since it is practically our only tool. The help that semantics can give us is minimal. For instance, we have no way of really determining whether the *ere* of Aquit. *And-ere* and that of Aquit. *Ere-xonis* are the same morpheme, even if their segments fully coincide (pace Gorrochategui & Lakarra 2001). On one hand, it is dubious that we can accept an analysis for Aquitanian in which there was a morpheme or allomorph ending in *-d*, as we would have if we split *And-ere*, which, in addition, would imply two roots with V- and V-/-V respectively, in PB. On the other hand, we apparently have two homophonous elements here, one in the first-member position, the other in the second-member position. This does not allow us to consider resolved the question as to the identification of the morphemes. Alongside this, the possibility of an eventual borrowing from Celtic remains open.

21 He is closest to such an approach in the fourth chapter on diphthongs added to the 2nd edition (of 1977), where he considers several processes at the syllabic level. Cf. now Jauregi (2007) on the synchrony and diachrony of the syllable in Basque.

In contrast to Lakarra (2002), I now believe that if *gizon* 'man', *gihar* 'lean meat', *gizen* 'fatty meat', *gibel* 'liver', and *sakon* 'deep', *sabel* 'abdomen', *samur* 'tender', *samin* 'strong pain', etc., are to be analyzed as **gi-zon*, **gi-har*, **gi-zen*, **gi-bel*, and **sa-kon*, **sa-bel*, **sa-bur*, **sa-bin*, etc. (cf. §5.11.2), then such formations cannot be considered to be composed of the two roots **gi-* and **sa-*, combined with their respective second members. In fact, **sa-* and **gi-* (as well as **la-*, **le-*, **za-*, etc.) do not have the necessary structure for every PB. root lexeme (CVC), nor the autonomy of other roots (they are not attested as independent words, nor to the right of any other known root). Hence, they can only be prefixes or old prepositions, unless we accept the dropping of root-final consonants, inexistent before the Middle Ages and, in any case, later than Aquitanian (cf. Michelena 1985 [1954b], Gorrochategui 1984). To sum up, these elements are prefixes, unless we cheat ourselves and change the definitions of *root* and *prefix* or, alternatively, ignore the chronology of the standard phonetic changes.

In several works (i.a. Lakarra 2004b, 2008c), I have systematically compared the bisyllabic root patterns, both between each other, and to the monosyllabic CVC roots. For this purpose, I have classified the words listed in Azkue's dictionary[22] into five groups, depending on their origin:[23] 1) loan words; 2) secondary variants; 3) compounds and derived elements; 4) phonosymbolisms; and 5) of unknown etymology.

Subsequently, I have applied a number of phonological and dialectological filters —which I consider to be rather lax (cf. Lakarra 2008c)— to the words classified in (5) above, in order to establish their level of adequacy as candidates for old roots. It is self-evident that (5), the list of roots of unknown origin, can only diminish in number throughout this filtering process, and that many words put there at the beginning will end up in (1), to a lesser extent in (3) or, occasionally, in (2) or (4), as I have been able to ascertain in several works from 2000 onwards. Thus, for example, the results of oCV (apparently the most im-

22 I have excluded those containing a diphthong because I believe that they are late (cf. §5.2 and now Lakarra 2011f), and hence they are a sign of compounding, derivation, borrowing, or secondary development due to a phonetic change (several examples are provided in §5.11.2). Consequently, they would make the statistical analysis unnecessarily difficult. In any case, they require further research.

23 At first, I started from the only consensus existent in Basque etymology, i.e. Michelena's work. The bulk of this is compiled in Arbelaiz (1978). The *FHV* itself, with its index, contains a great amount of etymological material. In addition, we now have Michelena's complete works at our disposal (cf. Lakarra & Ruiz-Arzalluz (eds.) 2011), with extensive word indexes. As further background reading, we also have Agud & Tovar (1988-1993), Corominas & Pascual (1980-1991), Rey (2006), and Palay (1991).

portant ones among bisyllables, and the closest ones to those of CVC; cf. §5.11.1) easily decrease to less than half if we take new advances into account: *oro* 'all', *oso* 'very', *otso* 'wolf' can now be explained internally as derivatives (Lakarra 2013), just as *olha* 'hut, forge', *alhatu* 'to feed, to make graze', *olho* 'oats', *orhe* 'dough' are to be explained as Gascon loan words (Lakarra 2011c, d, and here §5.9, in footnote). On the other hand, the application of phonotactic and geographical criteria has led —and will continue to lead— to the various groups of the remaining bisyllables becoming diminished to a minimal number (cf. Lakarra 2008c). In fact, those with a phonotactic structure which is compatible with Michelena's reconstruction (absence of affricates and initial voiceless stops, *muta cum liquida*, palatals, rising diphthongs, etc.) can only occasionally pass the dialectological criteria. In other words, there are very few bisyllables of unknown etymology, and of a phonotactic structure apparently compatible with that of Michelena's PB, which are attested in two or more non-contiguous dialects.[24] This is the case, even though the degree of attestation demanded is minimal (a single village, or even a single attestation in each territory, is enough to validate a word for comparison).

The percentage differences of "fossils" (words of unknown etymology which have passed through all the filters) are few, and probably not very significant among the diverse bisyllabic root patterns, but huge between those bisyllabic patterns and the monosyllabic CVC. Furthermore, these latter differences are ceaselessly increasing, insofar as many words of (5) —the great majority of which were within the bisyllabic patterns, in contrast to an extremely low rate of CVC elements— are relocated into some of the remaining groups (cf. §5.11.1, Table H).

It must be stressed that, until now, only those roots which have been really autonomously attested have been incorporated into the analysis, whereas the proto-forms reconstructed by Michelena or by myself (**bar*, **bor*, **dar*, **dats*, **dol*, **dul*, **ger*, **nin*, **non*, **zen*, **zon*…), or those which are obvious to every

24 I employ Bonaparte's classification (8 dialects), and not Zuazo's one (6 dialects), so as not to create additional difficulties when obtaining bisyllabic candidates for the PB. lexicon. However, "interfrontier" combinations (such as HN-L, HN-LN, HN-S), are accepted as valid, even if we know full well that the influence of the French-Spanish frontier was very late, and irrelevant for both PB and CB. Furthermore, the laxity of these filters goes as far as accepting combinations with very minoritary or unique attestations, such as 'Bermeo-Hondarribia', where the risk of these being loan words is obviously very high (since both cities are on the coast, although far away from one another). It is clear that, in the future, a more restrictive procedure will have to be employed. My prediction is that it will not affect monosyllables to any considerable degree, and will diminish the number of bisyllabic candidates to PB. (actually CB.) lexemes even more (cf. Lakarra 2011f).

analyst but which are not autonomously attested (such as **thor* 'to come', **khus* 'to see', *-bel* 'black', etc.) have been left out. It is clear that if my analysis were to be revised in view of these reconstructions, even adopting the most conservative procedure, the number of CVC roots would approximately double (practically all meet the phonotactic criteria, and are attested in the whole of the territory), whereas the bisyllabic survivors would barely increase in number (already currently low), faced with the strong reductions brought about by the increase of monosyllables.

As far as loan words are concerned, if we compare the numbers of monosyllables and bisyllables, we find that these are also very different. Loan words are very scarce among CVC elements, but reach up to 50% and 60% in patterns such as TVRV, TVSV, mVCV, RVTV, and RVSV (cf. §5.11.1, Table M).[25] By definition, the previously mentioned increase of the monosyllabic corpus from the reconstructed words would cause its percentage of loan words to decrease by half.

Finally, if we look at the right edge of root patterns, it is those ending in -C which appear to have the lowest rate of loan words and the highest rate of fossils —i.e. the most promising for housing old roots—, especially those with CV- (*la-*, for instance) to the left.[26] This supports my hypotheses about the CVC structure of the root, as well as the existence of prefixes and reduplications to the left in PB.[27]

---

25 Indeed, if we accept that most dialects are later than the 6th-8th centuries (cf. Michelena 1987 [1981a], Lakarra 2011f), the dialectal filter does not guarantee that we face candidates for PB. roots, but candidates for (at most) CB. Let this be interpreted with respect to both the extreme scarcity of bisyllabic roots and the notable abundance of monosyllabic roots for CB.

26 In c) of §5.8 (lVCVC), we would now have to considerably reduce those of unknown etymology. Some would clearly be loan words (*lehertu*, *lehor*, *lizun* < Lat. *tremere*, *termō*, *densu*), but many others would be derivatives from the prefix *la-*: *labar* 'cliff', *labur* 'short', *lagin* 'part', *lagun* 'mate', etc.; derivatives of other bases still remain, such as *lehen* 'before' (< **den-en*) and *lizar* 'ash tree' (in fact, the old variant is *leizar*; cf. *hareitz / haritz* 'oak tree'), and some derivatives such as *lotin* 'stody' and *lurrin* 'steam'. Also nVCVC could significantly decrease with respect to words of unknown etymology: *nafar* 'Navarrese' and *nabar* 'multicolored' are most probably patrimonial compounds, and *negar* 'crying' and (RS) *noxal* 'cow which has given birth twice' are loan words, as is *negu* 'winter' (cf. Agud & Tovar 1988-93) of nVCV. The % of fossils of all patterns significantly decreases (cf. §5.11.1). Cf. also the beginning of §5.9.

27 We should take into account, however, that many bisyllabic root patterns —not included in §5.11.1, as noted there— do not reach 10% in the attested words / possible words relation in the corresponding root patterns, so that it appears that they have not succeeded as actual patterns. The extraordinary abundance of patterns in -V in those cases appears to

5.6. From the canonical form to diachronic holistic typology

A monosyllabic root with prefixes and reduplications is not precisely the type of canonical form to be expected in a language like Modern Basque, which has universally been described as a quasi-perfect representative of agglutinative languages (cf. Trask 1998): SOV order, abundance of suffixes, copious verbal agreement markers, etc. Thus, for instance, in Proto-Uralic nearly 100% of the reconstructed lexical roots are CV(C)CV (cf. Bakrò-Nagy 1992), with the sole known and expected exceptions of grammatical particles and phonosymbolisms. The same holds true for Turkic and the Australian languages (Dixon 2002). In PB, and even later, the state of affairs was very different, if not diametrically opposite. By contrast, in the eastern monosyllabic languages —which are presumably similar to the oldest PB in this aspect—, there is hardly any inflection, their order is VO (and NAdj; cf. infra Mon-Khmer), etc.

Trask (1977), Gómez (1994) and Gómez & Sainz (1995) had already provided good reasons to believe that V had not always been final, but medial (Trask), or initial (Gómez, Gómez & Sainz) (cf. §10.4).[28] Greenberg (1963) himself noted that the NAdj order in Basque was, in principle, incoherent in a SOV language. However, he did not consider this to be an odd feature, since, in many other languages, this parameter is not harmonic with the unmarked order of constituents either, as e.g. in Modern English (AdjN, but SVO).

This idea that the word order NAdj within an OV typology is not at all unusual (and hence requires no diachronic elucidation), although perhaps satisfactory for some synchronic Bascologists, is actually a mistake, and proves to be incontrovertibly unacceptable from the diachronic point of view. At least since Meillet, we know that it is precisely the synchronic irregularities which must be exploited in diachronic research, since they are the elements with the greatest chance of continuing older states of the language. By contrast, in regularities, analogy and other innovating processes have erased the diachronically most relevant information. In the case of Modern English, the AdjN order is not coherent or harmonic with the current VO order, but it is with the earlier OV, which we positively know —as did Greenberg— the old language once had, like other

---

argue in favor of the (late) tendency towards the open syllable, typical of agglutinative languages.

28 Holmer (1970), hardly considered in later literature, classified the Basque language typologically with a number of Caucasian and Amerindian languages, but in a different group from that of agglutinative ones (cf. Lakarra 2011a). On the other hand, he assigns it accent on the second syllable, small roots, prefixes, phonological oppositions other than voiceless/voiced, etc. Cf. also Houis (1970) for a succint (and very similar) holistic typology of African languages.

Indo-European languages. Hence, the correct procedure in Basque should not be to do away with the NAdj order, and assign an SOV nature to the language from the beginning of time, but to search for other features coherent with a VO (and NAdj) order in the history or the prehistory of Basque.

Such features happen to exist and are attested in the history of Basque (although as a *fase sparita* in general), or can be reconstructed with certainty for earlier stages of the language (cf. Lakarra 2006b). Some of these features are:

a) AUX to the left of V (*da-tor* '(s)he is coming', *da-kar* '(s)he is bringing', etc.).
b) Analytic features.
c) Final accent (of both word and sentence) (cf. §4.6.4).
d) Prefixes and no suffixes.
e) Relative clauses to the right, not to the left, of phrasal heads.
f) Locative and possessive genitives to the right.
g) Inessive employed as "general locative", with no bi-univocal relationship between the markers of the diverse different local cases and their function (as in agglutinative languages).
h) Some others.[29]

In Lakarra (2005), (2006b), and subsequent works, I have suggested the existence of a profound drift in the early history of Basque, whereby it would have moved from a structure similar to that of the Mon-Khmer languages to the historical one, much closer to the structure observable in the modern Munda languages. In fact, both language groups are related (they belong to one of the two primary branches of the Austro-Asiatic family), but the structure of Munda is historically derived from another one similar to that of Mon-Khmer, which is thus archaic. Each structure corresponds to a group of features, which are represented in Table 5.2.

All this evidence argues in favor of the existence of an older language type, very far from the modern structure agglutinative / polysyllabic / ergative / SOV / pluripersonal inextricable V (pace Trask 1998). Obviously, not all of the arguments presented, or the different evolutions and the features discussed, will have the same value. As a matter of fact, research into most of these is nearly inexistent, so that it is probable and desirable that in the near future they will receive

29 One feature worth mentioning here is the archaic coordinative structure based on the sociative, as in RS. (246) *Ogi-agaz ur-a, or-agaz ero-en elikatura* bread-SOC.SG water-ABS.SG, dog-SOC.SG fool-GEN.SG food 'Bread and water, (~ they are) food of dogs and fools'. Stassen (2000) studies such structures, and divides languages into "WITH languages ~ AND languages", associating the former ones with the impersonal verb, the lack of inflection in NPs and some other analytic features in general. Cf. Lakarra (1983, 2008b), and §6.5.2.1.

different explanations, even opposed to the ones defended here. Even those arguments which keep their value will require important clarifications, and certainly not all arguments will be independent of each other.[30] This interdependence of all features is the essential point defended here. In other words, a holistic and comparative approach is the only approach which can help to present a new reconstruction of the pre-Michelenian PB, as well as to give a coherent explanation of the different features of the system and of the proposed changes. In this way, we will be able to go from the reconstructed proto-language to the real linguistic situation of the historical periods.

Table 5.2. Typological features of Munda and Mon-Khmer

	Munda	**Mon-Khmer**
Phrase Accent	Falling (Initial)	Rising (Final)
Word Order	Variable - SOV, AN, Postpositional	Rigid – SVO, NA, Prepositional
Syntax	Case, Verb Agreement[31]	Analytic
Word Canon	Trochaic, Dactylic	Iambic, Monosyllabic
Morphology	Agglutinative, Suffixing, Polysynthetic	Fusional, Prefixing or Isolating
Timing	Isosyllabic, Isomoric	Isoaccentual
Syllable Canon	(C)V(C)	(C)V or (C)(C)´V(C)(C)
Consonantism	Stable, Geminate Clusters	Shifting, Tonogenetic, Non-Geminate Clusters
Tone/Register	Level Tone (Korku only)	Contour Tones/Register
Vocalism	Stable, Monophthongal, Harmonic	Shifting, Diphthongal, Reductive
[Donegan & Stampe (1983: 337; 2004: 3, 16)][32]		

30 This is actually to be expected, if we do not search for synchronic or diachronic anecdotes, but for a solid basis of a general theory of the evolution of the Basque language.

31 Greenberg (1963) stated: "If in a language the verb follows both the nominal subject and nominal object as the dominant order, the language almost always has a case system". This is completed by Donegan & Stampe (2004) in the following way: "And if not a case system, we could add, then subject and object marking on the verb, as in Munda. There is rarely such marking either of nouns or verbs in Mon-Khmer and other languages of South-East Asia, where the verb precedes the object".

32 The same holistic pattern has been employed by Post (2011) to explain, without resorting to alleged loan words and substrates, the differences between the traditionally named *Indosphere* and *Sinosphere* languages.

5.7. The canonical form and the reconstruction of the Basque verb

Until recently, the monosyllabic canonical form has been studied only with regard to the NP. A number of well-known old verbs present this structure, but some others do not. That is why Lafon (1980 [1944]) discusses the verbal root structure in the following terms:

> "Pour avoir du verbe basque une connaissance satisfaisante, il faudra joindre à la connaissance de son système celle des rapports de forme, de sens ou d'emploi que certaines racines ont pu avoir entre elles, c'est-à-dire la connaissance de l'histoire et de la préhistoire des racines" (Lafon 1980 [1944]: I, 433).
>
> "B. *Racines simples*. Les racines simples du basque se composent d'une ou de deux syllabes. Les types suivants se rencontrent:
>
> 1º RACINES MONOSYLLABIQUES :
> a) une voyelle : *a-* 'être';
> b) consonne et voyelle : *go-* 'rester', *gi-* 'être fait';
> c) voyelle et consonne : *bil-* 'marcher', *kus-* 'être vu'.
>
> 2º RACINES DISSYLLABIQUES : *oa-* 'aller'; *augi-* 'venir', *aki-* 'être su', *zagu-* 'être connu'" (Lafon 1980 [1944]: I, 421).

In several earlier works (such as Lakarra 2006a, c), I have shown that this alleged polymorphism can and must be drastically reduced, and that the reasons for this apparent variation in the canonical form can be specified. Some of these are phonological (the dropping out of intervocalic consonants, and subsequent rise of diphthongs), and others are morphological (the attaching of one or several prefixes, in addition to the initial vowel). But the primitive canonical form which must be assumed for nouns and verbs is also CVC.[33] Table 5.3 shows several examples of the historical form that some of these CVC verbal roots have acquired and their origin.

In a considerable number of cases, the CVC structure can be reached by assuming that the original root was *dVC, and that *-d-* dropped out between vowels: **dan*, **daR*, **dats*, **den*, **din*, **don*, **dul*, **dun*, **duR*, **duz*, etc. We will

33 Such a conclusion is coherent with the verbal origin of many adjectives (cf. Post 2006, on the Tani languages), which have moved from being a subclass —with a very limited number of items— to being an independent category, progressively closer to nouns. As Gómez (1996) recalls, Humboldt pointed out that in Basque it was difficult to differentiate between Ns and Adj.s. Michelena (1988 [1964b]) attributes the property of being categorically less defined —even in later Basque— than in other languages, to the oldest strata of the lexicon: *hotz* 'cold', 'coldly' (N, Adj, Adv), *gose* 'hunger', 'hungry', 'with hunger', etc. Nonetheless, it should be noted that there are both monosyllabic and bisyllabic words among them.

come across this "imperfective" prefix (**da-*), which Trask (1977) had isolated in the finite forms of the present, later on in the chapter.

Table 5.3. Historical forms acquired by CVC verbal roots word-initially

Historical form	**Phonetic development from the canonical CVC**
VC	*utzi* 'to leave' < *eutzi* (Old G) < **e-duz-i*; cf. *luze* 'long'
CVC	*e-thoR-i* 'to come', *e-khaR-i* 'to bring', *i-bil-i* 'to walk', *i-khus-i* 'to see'
jVn	*jan* < **e-da-dan* 'to eat', *jin* < **e-din* 'to come, to become'
jVVn	*joan* 'to go' < **e-da-non* (> **edahon* > **edohan* (with usual metathesis *a-o* > *o-a*) > *johan*)
jaCVC	*jakin* 'to know' < **e-da-khin*
jaVC	*jausi* 'to fall' < **e-da-duts-i* (< **e-da-dun-tz-i*)

That this prefix was different from the causative *ra-* is evident on both formal and semantic grounds. Formally, neither *-VdV-* > *-VrV-* nor *-VrV-* > *-VdV-* can be adduced as regular phonetic changes in Old Basque. Semantically, the verbs in *e-dV- > *jV-* and those in *e-ra-* have different semantics in several of the clearest cases. Table 5.4 shows three basic *e-CVC verbal roots reconstructible for PB and their derivations with *-da-* and with *-ra-*.

*Table 5.4. *e-CVC verbal roots and their *e-da-CVC and *e-ra-CVC derivations*

PB HB[34]	**e-non* > *eho* 'to grind'	**e-da-non* > *johan* 'to go'	**e-ra-non* > *erho* 'to kill'
PB HB	**e-din* > *jin* 'to come'	**e-da-din-o* *jaio* 'to be born'	**e-ra-din* *eriden* 'to find'
PB HB	**e-san* > *esan* 'to say'	**e-da-san* *jasan* 'to suffer'	**e-ra-san* > *erasan* 'to attack'

Furthermore, we can also find combinations of both *da-* and *ra-* (and another residual one, *i-*), as shown in Table 5.5.[35]

To establish a single canonical formula, *e(-i)(-CV)(-CV)-CVC, as the verbal root, has important consequences for the PB. verbal morphology. To begin with, we reduce —as was desirable— the possibilities for variation, making the history of the verb more consistent, without having to wait *ad calendas graecas* until the comparisons with Caucasian provide some results. Contrary to Lafon's

34 HB = Historical Basque, from 1400 onwards.

35 On **dar* 'to sit', cf. Lakarra (2008a), and the therein cited bibliography, e.g. Aikhenvald (2006) or Frajzyngier (2002). On the origin —necessarily very different, in spite of the phonetic resemblance— of the causative *-ra-* (< **e-lo(n)* 'to take'; cf. *idoki* 'to remove'), and some related grammaticalizations, cf. Heine & Kuteva (2002) and Lakarra (2013).

expectations (cf. §5.3), these have proven unsatisfactory (apart from being unnecessary). Secondly, this procedure gives rise to a long series of verbal nets. In these, the root CVC is shared, and receives from 1 to 3 prefixes. Thirdly, the existence of morphology to the left of the lexical root becomes even clearer, a fact which is incompatible with the antiquity of the SOV order.[36]

*Table 5.5. *e-CVC roots, *e-da-CVC, *e-ra-CVC and *e-da-ra-CVC derivations*

PB HB	**e-zaR-i* *> ezarri* 'to put'	**e-da-zaR-i* *> jazarri* 'to attack'	**e-ra-zaR-i* *> erazarri* 'to admonish'	**e-da-ra-zaR-i* *> *jarzaRi > jasarri* 'to sit'
PB HB	**e-non-i* *> eho* 'to grind/kill'	**e-da-non-i* *> joan* 'to go'	**e-ra-non-i* *> erhan*, *erho* 'to kill'	**e-da-ra-non-i* *> eroan*, *eraman* 'to carry'
PB HB	**e-din-i* *> jin* 'to come'	**e-da-din-i* *> *jadin*, *jaio* 'to be born'	**e-ra-din-i* *> eriden* 'to find'	**e-da-ra-din-i* *> jarrain* 'to continue'
PB HB	**e-gotz-i* *> egotzi* 'to push'	--- ---	**e-ra-gotz-i* *> eragotzi* 'to prevent'	**e-da-ra-gotz-i* *> urgatzi* 'to assist'

On the other hand, if we count each of these verbal roots as a single case of conjugated root (with 3 or 4 possible radical extensions to the left), it is clear that Proto-Basque must have had no more than 20 or 25 synthetic verbs, and not the hundreds —or even the universal synthetic verb— postulated by many Bascologists, based on the reduction of these verbs from around 70 to fewer than 20 during the last five centuries (cf. Mounole 2011).[37] It is very likely, then, that PB was a language with a very limited verbal class, like many Australian and Neo-Guinean languages (cf. Pawley 2006), or some Amerindian ones (cf. Garrett 2005), and that the rise of the different Basque periphrases —at least the oldest ones, such as those consisting of ROOT + **edin/*ezan* or ROOT-*te*-INES +

36 As is well known, the fossilization of the synthetic causative is prehistoric (*ra-* → *-erazi*, *-erazo*), and apparently the applicative/directional (**da-*) must have fossilized at an even older period. The combination of both —always in the order **da-ra-*, not ***ra-da*— could possibly be older. The abundant irregularities found in the phonetic results of both are not at all surprising, quite the opposite, if we consider that such systematic irregularities are also found in the Bantu languages (cf. Good 2005). However, in these languages, such a phenomenon is synchronically productive, contrary to what has been taking place in Basque for a long time.

37 Incidentally, let it be recalled that we can find, among these, *joan* 'to go' / *eroan* 'to carry', *ibili* 'to walk' / *erabili* 'to use', *ekarri* 'to bring' / *erakarri* 'to attract', etc., which should count as one, according to the criterion defended here.

izan— was not always due to the Latin-Romance influence, as has been traditionallly defended (cf. Lakarra 2006a, Mounole 2006, 2007, etc., vs. Michelena 1987 [1981a], Gómez & Sainz 1995).

The origin and evolution of the root extensions[38] had not been studied before the development of the theory of the root, and it cannot be considered a settled issue.[39] In fact, it seems that, as in most monosyllabic languages, there must have existed, in PB, asymmetrical serial verbs (cf. Aikhenvald 2006), which later gave rise to verbal complexes, periphrases, and different grammatical markers in processes that have multiple parallels in other languages. Figure 5.1 shows some lexical sources for two TAM markers and for the verbal root *-i(n)-* (< ***nin*) which have been proposed in Lakarra (2006a) and later works (cf. §9.7.2 and §6.4.1.3, for further details about these grammaticalization processes).

I also believe that the origin of some elements outside the verbal morphology may be found in the grammaticalization of the hierarchically secondary components of the serial verbs. Some good candidates are the disjunctive particle *edo* 'or',[40] the 3rd grade demonstrative *har-* (probably from the verbal root *har* 'to take'; cf. Lord 1993) —whence the corresponding article—[41] or the for-

38 Cf. Hyman (2007) for Bantu, and Frajzyngier (2002) and Schuh (1977, 2003) for Cushitic.

39 In the old verbs *ihardun* 'to be doing something', *ihardetsi* 'to answer', *iharduki* 'to quarrel, to talk', or *jaregin* 'to release', which Michelena (1987 [1977]) explained as verbal compounds, we do not face an obscure or unknown nominal ***jar-* + light verb, but old verbs of historically known roots, but with a prefixing morphology. These are totally opaque and unexpected if we start from the traditionally adopted panchronic view of a morphology exclusively spread to the right. The same reasoning holds for verbs with a different word-initial form but which must start from the same morphological structure, such as *inarrosi* 'to shake, to make noise', *urgatzi* 'to assist', *jarrain* 'to continue', *ardan-* '*to drink' > 'wine', *arran-* 'to pull off, to take away' (cf. *arrano* 'eagle', *arrain* 'fish'), *arbin* 'to tighten', *arrats* '*to make descend' > 'nightfall', *jarrugi* 'to find', *jarrekin* 'to find', *eroan* 'to carry', etc.
In Michelena's *FHV*, we can read that, loan words left aside, *j-* only appears as a successor of the prefix *e-* of old verbs. Hence, *jar-* and its variants *ar-*, *inar-* and *ur-* cannot be nominals —in fact, they do not appear as such anywhere else in the morphology— but the result of the fossilization and amalgamation of the old prefixes of such verbs.

40 In my opinion, < **e-don* 'verbal prefix-to take'. For the loss of *-n* > *-Ø*, cf. *ido(n)-ki* 'to take out', *jo* 'to hit', *lo* 'to sleep' (from the same root, but with the expected change *d-* > *l-*), *ediro* 'to find', *eramo* 'to carry', *igaro* 'to pass', *ero* 'to kill', etc. On the origin of *jo*, cf. Ariztimuño (2012a).

41 **-har* > *-ha* (still attested as such in 1025; cf. Manterola 2006) > *-a*. The B. dialect still maintains *gizoná* 'the man' vs. *gizón á* 'that man there'. On the other hand, if we accept **-r* > **-VhV-* (*zur* 'wood' but *zuhaitz* 'tree', *hur* 'water' but *uhalde* 'river', *hor* 'dog' but

mation of equatives and comparatives with 'to go' / 'to come' and 'to pass', as in *adin* 'age, coetaneous; understanding', *adina* 'as much', < **har-din* 'that there-to become',[42] *-ago* 'more than', < **har-go(n)* 'that there-to cross'. Likewise, the noun declensional suffix *-raino* 'up to' might come from **ra-din-no* 'allative-to become-up to' (cf. §6.5.1.2), and *-rantz/-rontz/-runtz* 'towards' from **-ra-don-tz-i* 'allative-to take-old verbal suffix-participle' (cf. Gómez 2005, and §6.5.1.1). More on this can be found in Lakarra (2006a) and (2008a).

Morpheme in HB	**Lexical source**	**Typological parallels**
da- (ipfv.)	**daR* (cf. *jarri* 'to put' < **e-daR-i*	I.a., several Tibeto-Burman, Papuan, Australian languages (cf. Newman (ed.) 2002)
-en (pst.), *-de* (3rd p. pl.)	**den* 'to finish' (cf. *edeki* 'to remove', *etete-* 'breaking, finishing', *lehen* 'before')	I.a., Olutec (cf. Zavala 2006), Kristang (Portuguese-based creole) (cf. Baxter 1988), several Tibeto-Burman languages (cf. Solnit 2006)
-i(n)- 'to give', but also ethical dat., aux. of dat., *-i* (dat. marker in NPs)	**nin* 'to give'[43]	I.a., Ewe, Awutu (cf. Lord 1993: 39-41), several languages of New Caledonia (cf. Lichtenberk 1985), Saramaccan (cf. Byrne 1990)

Figure 5.1. Proposed lexical origins of da- *(ipfv.),* -en *(pst.),* -i(n)- *'to give' ~* -i *(dat.)*

5.8. Canonical form, word families, and the reconstruction of the Old PB phonology

> "The reconstruction of Old Chinese is sometimes treated as a pure matter of phonology. (...) The central thesis of the present book is that in order to reconstruct Old Chinese phonology, a proper understanding of Old Chinese morphology and word-families is indispensable" (Sagart 1999: 1).

When talking about the formation of the verbal system, I have pointed to the existence of multiple nets or families of words built upon the same verbal root and

ohalano 'mastiff'), the impossibility of **hC (*ozar* 'big dog' < **ohzar* < **hor* + *za(ha)r*), and **h...h...* > *Ø...h...*, then the postposition *-aga* (> *-ak*) can only be **har* + *-ga*.

42 Cf. Michelena (*FHV*). Obviously, this etymology renders any identification with the Iberian *ADIN-* impossible, in spite of the multiple attempts to relate them (cf. Lakarra 2010b and previously Michelena 1992, who was already conscious of such consequences).

43 As was indicated in Lakarra (2006a), only starting from **nin* is it possible to combine the common origin of *-n* and of *-i*, explained by Trask (1990), with the relative abundance of verbal roots ending with *-n*.

various simple or combined prefixes. Such word families are obviously not exclusive of the verb, but also spread across the NP. Table 5.6 shows some of these possible word families.[44]

Table 5.6. Word families derived from a single root

R	**Ø-R**	**e-R**	**2nd element**	**RED**[45]
**dan*	---	*edan*	---	---
**dar*	*lar*	*jarr(i)*	*in-dar*	*adar*
**dats*	*lats*	*jats(i)*	*al-dats*	*adats*
**datz*	*latz*	---	*ar-datz*	---
**den*	*lehen*	*edeki*	*ez-ten*	---
**der*	*ler*	---	---	*eder*
**din*	**lin*	*jin*	*gor-din*	---
**dol*	---	*jori*	*ozkol*	*odol*
**don*	*loh-i*	*jo / idoki*	*-do.i*	---
**dul*	---	*ilki*	**(h)ertz-dul*	---
**dun*	---	*e(d)un*	*bizar-dun*	---
**dur*	*lur*	*iurri*	*(h)an-dur*	---
**duz*	*luz(e)*	*(e)utz(i)*	---	---

As can be observed, there are alternations of the type *l-* (nude form) : *j-*/*d-* (prefixed forms) : *-d-* (2nd member of compound) : *Ø-* (reduplication), in the families corresponding to the Old PB. roots in *dVC, which had already become lVC in Standard PB. A general overview of the families allows us —as Michelena (*FHV*) predicted— to hit upon the destiny of (a great part of) the disappeared initial dentals (cf. Lakarra 2006a and, in §5.11.2, *jabe*, *jaio*, *jarein*).

On the other hand, this approach enables us to elaborate a kind of minimal grammar of the word formation in PB, based —as noted in §5.4— on reduplication and prefixing, and, perhaps later, on compounding.[46] Suffixation, much more widespread than prefixation and present in any language type (cf. Bybee & Pagliuca & Perkins 1990, Hyman 2008, Hawkins & Gilligan 1988), was probably also existent, although to a much lesser extent than in later periods.[47]

---

44 An interesting task for the future would be precisely that of researching the category of the reconstructed CVC roots.

45 To be read: R = Root; RED = Reduplication. I dispense with the meaning of words, since our interest in this analysis lies in the forms.

46 Here I differ from Gorrochategui & Lakarra (2001), where we presented a much more traditional view, rejecting the existence of prefixes and insisting on suffixing derivation and compounding as the really productive morphological mechanisms of PB.

47 In Lakarra (2006a), I have argued for an increase in suffixation in the recent historical period (cf. Sarasola 1997), as well as for the conversion into suffixes of some old prefix-

The study of lexical families also invites us to rethink the phonological system of Old PB, as Martínez-Areta (2006) did with far less etymological evidence than we have at our disposal now. If the language was monosyllabic at some point, then there would be no word-medial position, only a word-initial and a word-final position. This entails an evident need of reformulation of the Michelenian pattern (see §5.3).

In a period prior to that pattern, with no word-medial position, it seems that, in the realm of sibilants and sonorants, lenes and fortes were never opposed to each other. Hence, only a series of sibilants and sonorants existed, the lenes and fortes being in complementary distribution (lenes allophones appearing initially, fortes allophones appearing finally: (apico-alveolars) [ś-] ~ [-tś] / (dorso-alveolars) [s-] ~ [-ts]) (cf. Lakarra 2011b, e).

As for sonorants,[48] the only certain one in initial position seems to be *n*-, but only for the monosyllabic period (**nal*, **nan*, **nin-tz*, **nol*, etc.).

The historical cases of *l*- in Basque appear to have two origins: 1) old cases of **d*- (through the expected change **d*- > *l*-), and 2) loan words. Here are some phonotactic data of *l*- and of *n*- in the Basque historical lexicon:[49]

a) lVCV [92]:
   - Loan words: *labe*, *laga*, *laka*, *lako*, *lapa*, *laru*, *lasa*, *laso*, *lata*, *laza*, *lazo*, *lazu*, *leba*, *lega*, *lege*, *leka*, *leko*, *leku*, *lela*, *lena*, *lera*, *lerro*, *liga*, *ligu*, *lika*, *liki*, *lili*, *lino*, *lira*, *liro*, *lisa*, *lisu*, *loka*, *lona*, *lore*, *losa*, *lotsa*, *loza*, *luki*, *luku*, *lupi*, *lupu*, *lutu*, *lula*, *lutso* (45),
   - Unknown etymology: *leta*, *liba*, *liku*, *lope*, *loro*, *luba*, *lupa*, *lutho* (8).

b) lVCCV [53]:
   - Loan words: *lanbo*, *landa*, *langa*, *langi*, *lanpo*, *lantu*, *lantza*, *larde*, *larga*, *largo*, *lazka*, *lerdo*, *lerma*, *lertzo*, *lezka*, *lindo*, *lixta*, *lüzko* (18),
   - Unknown etymology: *lanbo*, *lanka*, *lenka*, *lisko*, *lisna*, *longa*, *longo*, *losko*, *lunka* (9).

c) lVCVC [66]:
   - Loan words: *laphitz*, *lapur*, *lasun*, *latil*, *legen*, *legez*, *legun*, *leher*, *lehor*, *lizun*, *likin*, *loben*, *lokan*, *lozan*, *lozer*, *lukur* (16),
   - Unknown etymology: *lahar*, *lapar*, *laputx*, *latsun*, *latzin*, *legatz*, *legar*, *likar*, *likits*, *lipar*, *lipitx*, *litsar*, *lokots*, *lorrin*, *lurrin*, *lutxar* (16).

---

es (**da*-, **de*-, **gi*-, **la*-, **le*-, **za*-), as has also happened in Munda (cf. Donegan & Stampe 1983, 2004) or in Australian languages (cf. *Diachronica* 2006).

48 Here (cf. Lakarra 2010a, c), I leave rhotics out of consideration, since they present specific problems, in both initial (actually word-initial rhotics do not exist in historical Basque) and final position (the neutralization appears to be later and more incomplete with rhotics than with the rest of the phonemes).

49 Again, I dispense with the meanings of words, since my treatment of them is only quantitative. The number in brackets indicates the total number of existing words in the attested lexicon corresponding to each pattern.

d) lVCCVC [30]:
- Loan words: *lanbas*, *lanjer*, *lanpas*, *lantes*, *linbur*, *lister* (6),
- Unknown etymology: *laztan*, *lerden*, *listor* (3).

a') nVCV [41]:
- Loan words: *naba*, *nabo*, *nadi*, *naga*, *nano*, *nasa*, *nata*, *nato*, *natu*, *natza*, *naza*, *nazi*, *negu*, *neka-1*, *neka-2*, *neke*, *nika*, *noku*, *nota*, *nunu* (20),
- Unknown etymology: *naka* (1).

b') nVCCV [9]:
- Loan words: *narda*, *neska* (2),
- Unknown etymology: *nazka* (1).

c') nVCVC [23]:
- Loan words: *nabal*, *nekos* (2),
- Unknown etymology: *nakan*, *nakar*, *negal*, *nekots*, *noxal* (5).

d') nVCCVC [7]:
- Loan words: — (0),
- Unknown etymology: — (0).[50]

As Michelena (*FHV*) already pointed out (at that time as a mere possibility still impossible to demonstrate), the origin of the medial fortes in one case and the other was to be searched for in consonant clusters. Some possible etymologies yielding word-medial fortis sonorants and sibilants are shown in Figure 5.2.

While the apico-alveolar /tś/ always seems to come from consonant clusters, the dorso-alveolar /ts/ appears in the final position of some old roots (*atz-o* 'yesterday', *etz-i* 'the day after tomorrow', etc.), and hence it would seem that it was due to a later phenomenon, like that of neutralization in favor of the fortes.

Within the sphere of rhotics, the only position in which the opposition historically exists is the intervocalic one, precisely the one which is missing in a phonology of the CVC type. The numerous minimal pairs alleged by Michelena in the *FHV* are questionable for an earlier phase in the majority of the cases, since they are next to a morphological barrier: *ere* 'also' / *erre* 'to burn', *ero* 'fool, to kill' / *erro* 'root', *hori* 'that; yellow' / *horri* 'to that; sheet', etc. We

50 Today, we ought to reduce the group (5), "of unknown etymology", and increase (1) "loan words", as well as (3) "compounds and derivatives" (the latter being particularly copious in lVCVC). In the parallel pattern in *d-* and above all *t-*, along with the abundance expected in loan words, the copiousness of phonosymbolisms is also remarkable: *danba*, *danda*, *danga*, *dinbi-danba*, *dindi*, *ddinddi*, *dinga-dinga*, *disti*, *dolda*, *dunda*, *dundu* (11), *daldal*, *dandan*, *dandar*, *dardal*, *dardas*, *denden*, *din-dan*, *dindil*, *din-din*, *ddinddin-ddanddan*, *dirdiz*, *dunbots*, *durdus* (13), *tada*, *tago*, *taka-taka*, *tala-tala*, *tana*, *tapa-tapa*, *tarra*, *tata*, *tato*, *tipi*, *tirri-tirri*, *tita*, *toto*, *tupa*, *tutu* (15), *takez*, *takitx*, *tarras*, *tetel*, *tilin*, *tirrin*, *totel*, *toton*, *tupotz*, *tutin* (10), *tango*, *tanpa*, *tanta*, *tindi*, *tunba*, *tunpa*, *tormu*, *torta*, *torto*, *tunka* (10), *taltal*, *tankan*, *tanketz*, *tantan*, *tantar*, *tartar*, *tertel*, *tiltil*, *tintin*, *tistil*, *tuntun* (11).

have now seen that many of these oppositions can be the result of (relatively) modern changes, such as *l* > *r* in intervocalic position (in particular after **e-* of the verbal prefix), or the rise of *-R-* out of consonant clusters, especially of /n-r/ with **d(a)* > *-n-* of an old verbal prefix: **e-da-ra-din* > **enaradin* > **eãrain* > **eanrain* > *jarrain* 'to continue'; and in a similar way in the old verbs which have jumped out of the verbal morphology to other lexical categories: *ar-rau(n/l)-tza* 'egg' (perhaps < **e-da-ra-don-tza*), *arbin* 'to tighten' (< **e-da-ra-bin*), **ardan-* 'wine' (< **dan* 'to drink'), *inaurri* 'ant' (< **dur* '(drag on the) ground'), *inarrosi* 'to shake' (< **non* 'to move, to stir'), *arrain* 'fish' / *arrano* 'eagle' (< **don* 'to put, to take out'), etc. Once the multiple rhotic *-R-* arose in this manner, it seems that the next step was its extension to final position and preconsonantal position, to the detriment of the old /r/.

fortis L:	**bel-le* 'crow'; **enon-le* > *ilhe* 'hair'; *Haran-luze* > *Haraluze* '(toponym) lit. long valley'
fortis N:	**arran-no* > *arrano* 'eagle', **ba-din-no* > *baino* 'but'; **-ra-din-no* > *-raino* 'up to'
fortis Ś:	**hor-tz-so* > *otso* 'wolf'; **her-tz(-te)* > *etse* 'house'; **e-ra-khus-tz* > *erakuts* 'to show'; **hartz-so* > *atso* 'old woman'

Figure 5.2. Some possible origins of sonorant and sibilant word-medial fortes

In view of the above, the emerging consonant system would be simpler than that of Michelena's PB: two sibilants (one apico-alveolar, the other one dorso-alveolar, both with fricative and affricate allophones, but with no contrastive opposition in their manner of articulation), three sonorants (*n*, *l*, *r*), and five stops (*b*, *t*, *d*, *k*, *g*, the phonetic nature of the phonation opposition between *t*, *k* and *d*, *g* still to be specified), in addition to the aspiration /h/. In fact, /h/ may have arisen from many other sounds (cf. Janhunen 2007, and obviously Michelena's *FHV*), as later occurred in the historical language. But, since there is no evidence in favor of that possibility for all words with a historically attested aspirate, /h/ must also be maintained for Old PB (cf. Lakarra 2011b).[51]

51 For various aspects of the evolution of *h* in post-Aquitanian Basque, and particularly for the etymological nature of many cases of *h* considered "adventitious" by Schuchardt and some others, cf. §5.2, and the bibliography cited in footnote.
In Lakarra (2011c) and (2011d), I have proposed three new etymological origins for multiple cases of *h*, to be added to Michelena's four origins:
(5) **-r-* > *-h* (*hur* 'water' > **huh-* > *uhalde* 'river' / *ibi* 'ford'),
(6) Gascon *h-* (*hami* 'hunger', *hein* 'size', *herabe* 'shy', *hobi* 'concavity', etc.), and
(7) metathesis of the type *hVR-* > *VRh-*, in both old loan words (*anhoa* 'daily ration' < **hanoa* < **aNoha* < Lat. *annōna*) and patrimonial words (*erhi* 'finger', *erho* 'fool, to

In the vowel system, we have not as yet found anything noteworthy, except for the confirmation of the (in some cases very) late emergence of all the diphthongs: *gau* 'night' perhaps < Lat. *cadō*, *haitz* 'rock' < **enaetz/*enaitz* (proposed by Michelena), *goi* 'up' < **gon-i*, **eurten* (> *i-/u-/erten*) 'to go out' < *egorten*, Old G. *eutzi* (> Old S. *eitzi*, central *utzi*, B. *etxi* > *itxi*) 'to leave', *artzain* 'shepherd' < **ardi-zani*, etc.[52] On typological grounds, however, the language type envisaged until now should perhaps lead us to expect a more copious vowel system, as in Proto-Munda (cf. Donegan 1993, Donegan & Stampe 2004), or Proto-Japanese (cf. Frellesvig & Whitman 2008).

The monosyllabic structure of the root, with prefixes, leads us to postulate a rising accent of both word and sentence, as in Mon-Khmer and contrarily to Munda, to continue with the case made famous by Donegan and Stampe.[53] As is well known, the old accent on the 2nd syllable is what Michelena postulated for late PB, although for quite a different reason (the distribution of the historical aspiration). The extreme weakness and obscurity of the initial position was noted —particularly with regard to stops— in the *FHV*. This feature does not invite us to locate the PB accent on that syllable (cf. Lakarra 2010a). In addition, we now have strong evidence for the presence of the radical monosyllable in the 2nd syllable, as well as of reduplications, prefixes, and hierarchically lower 1st members of compounds. This advises us, to an even greater extent, against an accent on the 1st syllable (see also Holmer's observation recalled in §5.6).

---

kill', *ulhain/unhain* 'drover'), or Gascon loan words (*elhe* 'to speak', *unhatu* 'to get tired', *orheit* 'to remember'), etc.

By contrast, some others appear to be segmentally unmotivated. Among these, we can include *sorho* 'meadow', *bulhar* 'breast', and the majority of those with a CVRh- structure.

52 Recall what was said (§5.3) about the structure of the historical phonosymbolisms and their consequences for research into the canonical form of the radical lexemes.

With regard to the causes of the alleged abundance of initial V-, I have discussed them several times since Lakarra (1996b, and now 2012a), where I denounced the fraudulent use of that and other phenomena made by Vennemann (1994) to justify his feeble "Proto-Vasconic" (cf. also Lakarra 2010a and 2011b).

53 Cf. Lakarra (2006b). Already in Lakarra (1996b), I proposed that the assumption of bisyllabic feet for classical PB, of a structure [_ '_], allowed us to combine Michelena's and Martinet's explanatory aims. The former aimed to account for the distribution of the aspiration, while the latter established a demarcative accent which could account for the autonomy of the word, and consequently for the different evolutions of voiceless stops —or, we should add, of Latin sibilants and sonorants— in initial and medial positions.

5.9. Changes in the canonical form

As was noted in §5.4, the radical patterns analyzed (see the Appendix in §5.11.1)[54] are divided into two groups with respect to almost every feature of antiquity: i.e. % of roots of unknown etymology, or the opposite % of loan words; % of attested roots over the possible combinations; % of roots which can get through the phonotactic and geographical filters, etc. On one hand, we have the radical pattern CVC, and on the other, the remaining ones (all bisyllabic). The differences between the 1st and 2nd groups are very relevant in all sections of the tables. As we have already pointed out, they can only significantly increase (up to around double) if we bring into the analysis —in addition to the roots autonomously attested in the historical periods— the roots obtained by means of reconstruction, however conservative this should be.

In fact, the figures corresponding to the "fossils" —i.e. the members of (5) which have passed through the phonotactic and geographical filters, and thus the only ones which have some chance of having been PB— are virtually negligible in bisyllabic patterns. If we take those which have between 1% and 7% of fossils (oCV 7.05%, aCVC 3.52%, oCCV 2.79%, aCV 2.35%, uCV 2.35%, aCCV 1.85%, SVTV 1.33% and oCVC 1.17%), and we look at the absolute data of these fossils, we find 6[55] / 15 / 6 / 2 / 4 / 4 / 5. Hence, providing the etymology of a dozen of these fossils —quite a feasible undertaking— would suffice to render these percentages totally irrelevant when reconstructing both the Old PB and the more recent one.

Furthermore, I have already pointed out that, although the phonotactic filters correspond to the "classical" PB (the last centuries previous to the Christian Era), the geographical ones —based on the dialectal division sketched by Bonaparte— can only correspond to a much later period, not only post-PB, but even later (by definition!) than Michelena's CB (5th-6th centuries).[56] And even for that

54 In the tables of the Appendix in §5.11.1, 20 root patterns are studied, although in some previous works (as in Lakarra 2009b) I analyzed 50. However, here I prefer to stick to those in which at least 10% of the potential roots are attested, as I said in §5.5 (in fn.).

55 Of which we should now eliminate *oro*, *oso*, *otso* (patrimonial derivatives), and *olho*, *orhe* (Gascon loan words) (cf. §5.5), as a result of which the percentage (7.05%) drops to less than 1.2% (cf. Lakarra 2011c, d). It is very probable that the rest of the patterns also fall to less than 1% if they are examined in more detail.

56 See Lakarra (2011f) for some implications of the bipartite model of dialectal differentiation. In any case, we would not have the totality of the dialects described by Bonaparte (8), Michelena (12), or Zuazo (6), until much later (than the 5th-6th centuries). We should significantly postpone the chronology to which the root patterns could be dated through those dialects not only with respect to the recent PB, but also with respect to the previously mentioned 5th-6th centuries.

period, very close to the 2nd millennium of the Christian Era and far from the recent PB, we cannot ensure the historical majority of bisyllables, nor even the existence of any significative bisyllabic pattern, exactly the opposite to what happens with CVC.

To reach the bisyllabism of the historical period, we can assume a drift similar to the one observed in Chinese (Feng 1998, Duanmu 1998), or in Tani (Post 2006). A number of phonological and morphological processes conspired to erase the old morphemic boundaries in favor of other, wider ones: vocalic metatheses (*hedoi* > *hodei* 'cloud'), or consonantal ones (**e-da-zun* > **ezadun* > *ezaun* > *ezagun* 'to know'),[57] dissimilations (**hark-har* > *elkar* 'each other') and assimilations (*zin-hets-i* > *sinetsi* 'to believe'), the elimination of similar consonants (*solas* 'conversation' < Rom. *solaz*, *sasoi* 'seasoning' < Sp. *sazón*, or *inardun* 'to be occupied in doing something' < **irardun*), metatheses of the type $*h_3/*h_2 > h_1$ (**e-lin-o* > *herio* '(active) death', **enuskara* > *heuskara* 'Basque language'), metatheses of the type *hVR-* > *VRh-* (**heri* > *erhi* 'finger', **hilun* > *ilhun* 'dark', Gasc. *hele* > *elhe* 'word'), dissimilations of aspirates (**hil-herri* > *ilherri* 'cemetery'), movements of *r* or *n* to the right (Lat. *tremere* > **derne* > *leher* 'to explode', **hatz-zarpa* > *atzapar* 'claw', Sp. *dengue* > *legen* 'leprosy', Lat. *offende(re)* > **obeden* > **obenen* > *hoben* 'offence, sin'), contractions of polysyllables (**e-da-din-o* > *jaio* 'to be born', **enazur* > *hezur* 'bone'), epentheses of the antihiatic *-g-* (**ha.in* > *hagin* 'tooth; yew tree', **ja.on* > *jagon* 'to look after'), the dropping or neutralization of the last segment in the first members of compounds (*ogi* 'bread' > *ot-*, *buru* 'head' > *bur-*, *baso* 'wood' > *basa-*), etc.

Along with this —and in fact, most probably long before—, the old nominal or verbal prefixes fossilized (**la-ban-i* > *labain* 'slippery', **sa-bur* > *samur* 'tender', *gi-bel* > *gibel* 'liver; rear part', **e-da-non* > *johan* 'to go', *e-ra-bil-i* > *erabili* 'to use (= causative of *ebili* 'to walk'), **e-da-ra-don* > *arran-*, etc.). Reduplications also ceased to be productive (*ze-zen* > *zezen* 'bull'). In some cases, the dropping of the onset of the first syllable blurred the old and originally reduplicative nature of the word, turning it into a mere puzzle for the diachronist (**da-dar* > *adar* 'horn, branch', **no-nol* > *ohol* 'plank'). In these three cases, the respective original roots were **zen*, **dar*, **nol*. There is, however, no trace of a reduplication starting from a lVC root, such as ***lu-lur* (reduplication of *lur* 'earth' < **dur*) and this is probably related to the fact that many historical *l-* of Basque have been **d-* at some previous phase (cf. *FHV*).[58] Finally, there also

57 On metathesis, cf. Egurtzegi (2011).

58 As for whether there are hVC roots in reduplications, *HAHAN-* is found in Aquitanian, perhaps from the root **han* (cf. *handi* 'big', *ahuntz* 'goat' < **han-huntz*, *aker* 'billy goat' < **han-ger,* etc.). Since **h* is postulated for Aquitanian and PB, there is no need to analyze *h-* as coming from **th-*. Hence, *hahan* could be as old as *zezen*, etc. I know of no

exist material proofs (cf. Lakarra 2006a), and not only formal evidence, that a change has taken place whereby some originally prefixing elements became suffixes, as is observed in §5.8 (in footnote).

While the morphological changes appear to be very old (earlier than Michelena's PB), the phonological ones are in general much later, and, in most cases, medieval. Because of all this, and the reasons given in §5.5 about the late nature —later than CB— of all the bisyllabic radical patterns, I have argued that the new dominant canonical form is recent —medieval— in the evolution of the history of Basque. Obviously, the drift towards the agglutinative type (SOV, bisyllabic, suffixing, etc.) would have begun much earlier, perhaps several millenia before, at least if Donegan & Stampe (1983, 2004) were right in assigning a longer developing process to this kind of drift than to the inverse one (observed e.g. in the Kwa languages; cf. Hyman 2004).

In any case (cf. Lakarra 2009b), for the dominant canonical form to be considered bisyllabic, not only is it necessary that many bisyllabic phonological forms should exist, but also that the majority of the inputs of the formation of new words should be bisyllabic. However, nearly all of them (*harbel* 'slate' < *har-bel* 'black stone', etc.) must have been transparent until a rather recent period, even though reduplications and suffixes had already been fossilized by Modern PB.

5.10. Some conclusions and remaining tasks

The analysis of the canonical form of morphemes and, in particular, the theory of the monosyllabic root, is the core of the reconstructive paradigm undertaken here. The results obtained here are not definitive, nor do they affect the whole of the PB. lexicon, phonology, morphology, and syntax (a pretension which would be excessive, probably futile, and counterproductive for the advance of the research). I believe, however, that we now know the prehistoric structure of the language better than several years ago. Thus, we are conscious of new generalizations, as well as of new aspects of an earlier grammar, such as reduplication and the use of prefixes on both verbs and nouns. In addition, we now know some new etymologies, and this etymologizing enterprise is expanding and consolidating, as new findings are at our disposal. At the same time, we have come to know, and partially to explain upon a more solid basis, multiple changes in

case of reduplication of hVC with *h-* < *th-*, but this hardly proves anything, since there are very few certain examples, if any at all, of **th-* > *h-*.

every area of the language, from the earliest reconstructible stages to the first sufficiently attested ones (Archaic Basque).

1) Contrary to the manner in which the Indo-European or Semitic traditions —and more recently the Austronesian, Uralic or Sinitic ones— have evolved over the last few decades, prior to 1995, Bascology had not systematically studied the canonical form of the root, nor had it worked out any theoretical framework which would increase our knowledge of the old morpho-phonology, and hence be useful for the reconstruction of PB.
2) Although Uhlenbeck and Michelena on several occasions argued for the bisyllabic canonical form for the old radicals —and in fact, the latter started from a radical polymorphism in his reconstructing procedure—, a number of restrictions have been pointed out in the forms of morphemes, which impel us to rule out radical polymorphism, or even bisyllabism, as original (i.e. early) features.
3) The theory of the monosyllabic root CVC proposed for Old PB in Lakarra (1995), and developed in a number of subsequent works, has proven to be as productive as analoguous reconstructing processes based on the canonical form, developed in diachronically more researched linguistic traditions, such as the Indo-European, the Semitic, the Uralic, the Austronesian, or the Sinitic. The main achievements can be summarized in the following points:
    a. Previously undetected new roots have been obtained, and the number of compounds and derivatives corresponding to already known roots has been increased. Likewise, new loan words have been detected, starting from phonotactic features which are incompatible with the canonical form.
    b. Important aspects of the PB grammar, later fossilized, have been reconstructed, such as certain nominal prefixes (*gi-*, *la-*, *sa-*) and reduplications.
    c. The elaboration of a root theory or of specific interrelated phonotactic constrictions entail a filter or additional diagnostic for a rigorous comparison between PB and any other language, to be added to the standard phonetic reconstruction of PB.
4) The reconstruction and etymological analysis based on the canonical form of the roots has fostered a formal, not atomist, etymologizing procedure. This undertaking was already present as a vague idea in Uhlenbeck (1947 [1942]), but has remained undeveloped since then. For that purpose, a set of criteria were presented, such as the percentage of words attested with respect to that of potential words in each radical pattern, the percentage of loan words with respect to attested and potential words and, in particular,

the percentage of fossils (words of unknown etymology which have got through the phonological and geographical filters) with respect to the potential ones. It is clear that, even if no formal criteria can provide a precise etymology for a given word, the profile sketched by all the previously mentioned criteria of the pattern corresponding to that word provides us with an incomparable benchmark, unattainable by any other means.

5) The study of the canonical form has led us to the study of the typological drift from a structure harmonic with the CVC monosyllabism (isolating/analytic, prefixing, VO order, rising accent, inexistent or scant agreement, etc.) to another, different one, present in the last centuries of historical Basque: bisyllabism, agglutination, suffixing, OV order, copious and increasing agreement, just as in the nominal inflection. Some of these features, however, are late, and features of the earlier stage survive, such as the NAdj and NRel orders.
6) The confirmation of the monosyllabic canonical form of the verbal root has multiple implications. To begin with, the existence of a unique radical pattern in the CVC form is opposed to the polymorphism defended by Lafon for verbal roots; from this CVC pattern, the different stems would derive through phonetic changes and morphological additions of earlier prefixes, all of them fossilized in the historical period. Secondly, we can verify the existence of various grammaticalizations of typical serial verbs, in a number of categories. Thirdly, we are led to believe that the class of synthetic verbal roots, although probably integrated in a dense network of forms derived through prefix combinations, was always very reduced, as in many Australian, New Guinean, or Amerindian languages.
7) The lexical families which are thus reconstructed around the old roots serve to provide a greater security in the etymologizing process —so far over 800 etymologies have been obtained—, and in the phonology of Old PB. I believe that it can be safely stated that the medial fortes sonorants of Standard PB come from old consonant clusters, and that in sibilants the structure with only two positions (initial and final) brings us to an earlier stage with a complementary distribution of fricatives and affricates. Thus, the number of sonorant and sibilant phonemes reconstructed for the earliest stage is halved.
8) The canonical form of morphemes is a long-lasting feature, and yet is subject to diachronic development. A variety of consonantal and vocalic changes at the edges of words and in medial positions have conspired to fossilize and blur the old morphematic (monosyllabic) barriers, thus creating new bisyllabic and polysyllabic structures. Likewise, new additional evidence has been found in favor of the late nature of

diphthongs, of the etymological value of multiple aspirations, as well as of three origins for the aspiration in addition to the ones postulated by Michelena, as well as in favor of the evolution of the roots with **d-* in pre-Michelenian PB, thus accounting for a number of grammatical fossils (**da-*, **la-*, **-da/-t*, **-la*, **-de/*-te/*-e*, **za-/*-za*, **-di/*-ti*) and over 50% of the potential lexicon of the above mentioned pattern *dVC.

9) Last but not least, I believe that we can say that the monosyllabic theory of the root in Old PB has been able to answer some old questions, to set others on the right track to a solution, and to show the weakness of several deep-rooted ideas about the structure and evolution of the history of the language. Furthermore, it poses, as can be expected from any new theory in the process of development, new questions concerning phonology, morphology and syntax, hitherto overlooked or considered irrelevant, turning them into issues for further research in the future.

I hope that it has become clear that, also in the field of Basque, there is a close relationship between reconstruction and etymology, linguistic theory and philology, scientific explanation and abstraction. We should aim at generalizations that go beyond the anecdotal, and etymological research should be carried out not (or not only) according to the etymologist's "idiosyncrasies", however curious and influential these might be, but as ineluctable results of the reconstructive models, and of the paradigms underlying that research.

I have also shown the real need —not just mere interest— to combine typology and traditional reconstruction, especially when this is internal and of 2nd grade (cf. Lakarra 2005), as it is in our case. Since the analysis of the canonical form of morphemes is the basis of the new paradigm, I consider it obvious that this path of research —however far it still is from having exhausted its potential— will require the help of non-genetic comparison (i.e. of diachronic typology) to reach a productive scenario of reconstruction, as was already aimed at in Lakarra (1998a). Thus, it will prevent our research from getting tangled up in the kind of petty details which had already exasperated Benveniste over seven decades ago. Let reconstruction and typology go hand in hand. This alone will enable us to reach real profundity in the reconstruction of PB, and a standard in the diachronic explanation of the Basque language which would have been unimaginable only a few years ago.

5.11. Appendices

5.11.1. Tables

I give now three types of tables which show the antiquity, structure and evolution of radical patterns. Table L shows attested and potential roots, whereas Table H shows the roots of unknown etymology and, in particular, those which pass the phonotactic and geographical criteria (= fossil roots). In Table M, loan words are studied; since this criterion is negative (the more loan words a radical pattern contains, the less probability it will have to be old), the table is organized in inverse order (from smaller to greater %).

I have considered it appropriate to gather only the radical patterns which are most significant to elaborate the chronology of radical patterns. Therefore, the tables include only those models which reach at least 10% of attested roots over the total amount of possible ones.

Notice also that only the monosyllables attested in isolation are included (for instance, *huts* 'empty', *sar(tu)* 'to get into; to take into', *gor* 'deaf' are included, but **bel* 'black', or unanimously accepted verbal roots such as *-thor-* 'to come', *-gin-* 'to make' are not). Hence, their percentages of "words of unknown etymology" (5) and "fossils" could only increase, in fact considerably as I see it, if criteria were relaxed and the most secure reconstructions were included. By contrast, the research about bisyllabic words without known etymology mostly leads them to being classified among (1) "loan words" or (3) "compounds and derivatives", or even (less frequently) among (2) "modern variants" or (4) "phonosymbolisms". Consequently, the rates given here are to be understood provisionally, and will probably be reduced in a significant way, as has already occurred on multiple occasions in a number of works since 2000.

Legenda: RP = Root Pattern; PR = Potential Root; A = Attested; UE = of Unknown Etymology; F = Fossils; LW = Loan Words; OPS = Onomatopoeia and Phono-Symbolisms; %-1 = F/LW/OPS divided by A; %-2 = F/LW/OPS divided by PR.

Table L

RP	PR	A	%	UE	%-1	%-2	F	%-1	%-2
aCV	85	64	**75.29**	04	06.66	04.70	02	03.12	02.35
uCV	85	44	**51.75**	05	11.36	05.83	02	04.54	02.35
CVC	325	152	**46.76**	70	46.05	21.53	48	31.18	14.76
aCCV	215	82	**38.13**	12	14.64	05.58	04	04.87	01.85
SVRV	200	71	**35.50**	05	07.04	02.50	01	01.40	00.50
oCV	85	29	**34.11**	07	24.13	08.23	06	20.68	07.05
TVRV	600	195	**32.50**	25	13.07	04.33	04	02.05	00.66
aCVC	425	110	**25.88**	33	30.00	07.76	15	13.63	03.52
SVTV	300	75	**25.00**	13	17.33	04.33	04	05.33	01.33
RVTV	300	69	**23.00**	09	13.04	03.00	01	01.44	00.03
mVCV	425	94	**22.10**	12	12.76	02.82	00	00.00	00.00
TVTV	900	190	**21.11**	16	08.42	01.77	00	00.00	00.00
uCCV	215	45	**20.93**	06	13.33	02.78	00	00.00	00.00
RVRV	200	41	**20.50**	02	04.87	01.00	00	00.00	00.00
oCCV	215	39	**18.13**	08	20.53	03.72	06	15.37	02.79
TVSV	600	95	**15.85**	14	14.73	02.33	03	03.15	00.50
CVmV	325	49	**15.07**	09	18.35	02.76	00	00.00	00.00
uCVC	425	54	**12.70**	05	09.25	01.17	01	01.85	00.23
hVCV	425	54	**12.47**	18	33.33	04.23	02	03.70	00.47
RVSV	200	24	**12.00**	00	00.00	00.00	00	00.00	00.00
oCVC	425	47	**11.05**	07	14.59	01.64	05	10.63	01.17

Table H

RP	**PR**	**A**	**%**	**UE**	**%-1**	**%-2**	**F**	**%-1**	**%-2**
CVC	325	152	46.76	70	46.05	21.53	48	31.18	**14.76**
oCV	85	29	34.11	07	24.13	08.23	06	20.68	**07.05**
aCVC	425	110	25.88	33	30.00	07.76	15	13.63	**03.52**
oCCV	215	39	18.13	08	20.53	03.72	06	15.37	**02.79**
aCV	85	64	75.29	04	06.66	04.70	02	03.12	**02.35**
uCV	85	44	51.75	05	11.36	05.83	02	04.54	**02.35**
aCCV	215	82	38.13	12	14.64	05.58	04	04.87	**01.85**
SVTV	300	75	25.00	13	17.33	04.33	04	05.33	**01.33**
oCVC	425	47	11.05	07	14.59	01.64	05	10.63	**01.17**
TVRV	600	195	32.50	25	13.07	04.33	04	02.05	**00.66**
TVSV	800	115	14.62	24	20.86	03.00	04	03.47	**00.50**
TVSV	600	95	15.85	14	14.73	02.33	03	03.15	**00.50**
SVRV	200	71	35.50	05	07.04	02.50	01	01.40	**00.50**
hVCV	425	54	12.47	18	33.33	04.23	02	03.70	**00.47**
uCVC	425	54	12.70	05	09.25	01.17	01	01.85	**00.23**
RVTV	300	69	23.00	09	13.04	03.00	01	01.44	**00.03**
mVCV	425	94	22.10	12	12.76	02.82	00	00.00	**00.00**
TVTV	900	190	21.11	16	08.42	01.77	00	00.00	**00.00**
uCCV	215	45	20.93	06	13.33	02.78	00	00.00	**00.00**
RVRV	200	41	20.50	02	04.87	01.00	00	00.00	**00.00**
CVmV	325	49	15.07	09	18.35	02.76	00	00.00	**00.00**
RVSV	200	24	12.00	00	00.00	00.00	00	00.00	**00.00**

Table M

RP	PR	A	%	LW	%-1	%-2	OPS	%-1	%-2
CVC	325	152	46.76	12	**07.89**	03.69	30	19.73	09.23
uCVC	425	54	12.70	05	**09.25**	01.17	01	01.85	00.23
oCVC	425	47	11.05	06	**12.76**	01.41	01	02.12	00.23
hVCV	425	54	12.47	07	**12.96**	01.64	04	09.25	00.94
aCVC	425	110	25.85	16	**14.54**	03.76	06	05.45	01.41
uCV	85	44	51.75	07	**15.90**	08.23	08	18.18	09.41
uCCV	215	45	20.93	08	**17.77**	03.72	03	06.66	01.39
aCCV	215	82	38.13	15	**18.30**	06.97	03	03.65	01.39
aCV	85	64	75.29	14	**21.87**	16.46	10	15.62	11.70
oCV	85	29	34.11	07	**24.13**	08.23	03	10.34	03.52
oCCV	215	39	18.13	10	**25.63**	04.65	00	00.00	00.00
RVRV	200	41	20.50	14	**34.14**	07.00	06	14.63	03.00
SVTV	300	75	25.00	28	**37.33**	09.33	03	04.00	01.00
SVRV	200	71	35.50	27	**38.02**	13.50	03	04.22	01.50
TVTV	900	190	21.11	80	**42.10**	08.88	36	18.94	04.00
CVmV	325	49	15.07	22	**44.88**	06.77	07	14.24	02.15
TVRV	600	195	32.50	95	**48.71**	15.83	16	08.20	02.66
TVSV	600	95	15.85	47	**49.47**	07.83	02	02.10	00.33
mVCV	425	94	22.10	47	**50.00**	11.05	16	17.02	03.76
RVTV	300	69	23.00	36	**52.17**	12.00	00	00.00	00.00
RVSV	200	24	12.00	15	**62.50**	07.50	00	00.00	00.00

5.11.2. Etymological appendix

As examples of the etymological work based on the analysis of the canonical form, I give below, briefly and without going into unnecessary details, 50 etymologies, gathered from recent works or works in course. Numbers in parentheses refer to words which are analyzed within the list of 50 (in such cases, their reference number is given).

1. ***adar*** 'horn, branch': **dar* → **da-dar*; cf. *gogor* 'hard', *zezen* (49) 'bull' for the reduplication, and *adats* 'hair' (< **da-dats*), *eder* 'beautiful' (< **de-der*), *odol* (39) 'blood', *ahal* 'can' (< **na-nal*), *ihin-tz* 'dew' (< **ni-nin-tz*), *ohol* 'plank' (< **no-nol*), etc., for the additional evolution *C-C* > *Ø-C*.
2. ***afari*** 'supper': **gau* 'night' + *hari* 'taken'. Cf. *ba(ra)zkari* 'morning snack' (< *baratze* 'vegetable plot'), *askari* 'mid-afternoon snack' (< *arrats* 'evening'), *bidari* 'traveling expenses' (< *bide* 'path') or *janhari* 'food' for *-hari*. Regarding the origins of /f/, cf. *ifar* (25) 'north' and *ifini* (26) 'to put'.
3. ****ardano*** 'wine': **e-da-ra-dan-o* (**dan* 'to drink'). This would occur after the dropping of the pretonic vowel (**edardano*) and the usual regular

change *-d- > -Ø-*, **ear- > ar-*, *-n- > -Ø-*; here there is no *-n- > -h-*, or at least not **-h- > h-* (***hardao*, etc.), presumably due to its occurrence in the 4th syllable, as opposed to *herio* (21) 'death' (cf. *jario* 'to spill', < **e-da-lin-o*) or *hodei* 'cloud' (but *jaio* (30) 'to be born' < **edadino*). Cf., also, *arrain* (4) and *arreba* (5), and some other nouns beginning with *aR-* from a former prefixed (verbal) amalgam **e-da-ra-*.

4. ***arrain*** 'fish': **e-da-ra-don-i* (**don* 'to put ~ to take'). Cf. **arraNo* 'eagle' < **e-da-ra-don-no*, both from **arran* (< **earnan*). Note the metathesis *a-o > o-a* (cf. *unhain* (45) 'oxherd', *urgatzi* (46) 'to assist', *joan* (33) 'to go'). For *aR- < *edara-*, cf. *arbin* 'to hold tight', **ardano* (3) 'wine', *arrau(n/l)tza* 'egg', etc.
5. ***arreba*** 'sister of a man': **e-da-ra-de(n)-ba*. Cf. *arrain* (4) and **ardano* (3), for the first element (*arr-*) and **den* (*lehen* 'before', *etete-* 'to finish, to break', *edeki* 'to remove') for the root. *Arreba* appears to be an antonym of *errain* 'daughter-in-law' (< **e-ra-din-i*), the latter coming from the causative of *jin* 'to come' (recall Benveniste's 1969: 239-243 reflection on the noun for 'woman' in Indo-European), the former coming from 'to remove, to take away'. Perhaps we had better say that both represent the two points of view of the same action (that of taking the wife to her husband's house): *errain* from the point of view of the husband's family (those who take the wife in), and *arreba* from the point of view of the wife's family (those who lose her, those she is "taken away" from).
6. ***atzamar*** 'finger'. It cannot come from **hatz-hamar* 'finger-ten', as proposed by Moguel and some others, as, leaving aside *bat* 'one' and *bi* 'two', numerals go before, not after the modifed noun. In fact, we have the compound *hamarratz* 'octopus', with *hamar* and *hatz* following the rules of word-compounding. One interesting variant which does not have to be necessarily late, *atzapar*, is perfectly derivable from the mixed compound **hatz-(z)arpa*: the *-r-* would have jumped from the coda of the first syllable to the coda of the second syllable, as in *dollor* and *leher*, etc. *Hatzamar* is probably a result of folk etymology, like *deihadar* 'clamor', in spite of Michelena's explanation of the latter as *dei* 'call' + *h* + *adar* 'horn' (cf. *FHV*). If we look it up in the *OEH*, it is evident that the meaning 'calling horn' is conceived by Orixe and some later followers, whereas in all varieties and periods of the language it means 'noise, clamor', which can be better explained if we start from the word family of *trueno* 'thunder, storm', with the same jump of /r/ as *atzamar*, etymological /h/ from **VnV*, and *Vh > Vih*, as in *aihotz* 'sickle'.
7. ***aulki*** 'chair': **abedulki* (i.e. Sp. *abedul* 'birch tree' + *-ki*, suff. indicating 'thing, object') > **abeulki* > **abulki* > *aulki*. Cf. the *-e* of *esne* 'milk' (<

**(b)eh(i)-sen(i)-e(d)a(b)e* 'cow'-'offspring'-'drink') for the dropping of intervocalic voiced stops, and *alu* 'vagina' (< Lat. *aluu*; Ernout & Meillet 2001 [1932]: 25: "ventre ou plutôt cavité intestinale […]. Se dit aussi pour *uterus*"), *hogen* 'fault' (< Lat. *offende(re)*, which later > **offeden*), *laur* 'four', *orein* (40) 'deer', etc., for *-b-* > *-Ø-*.

8. ***begi*** 'eye': **goi-hegi* 'upper part-corner' > **gwi-egi* > **biegi* > *begi*. Cf. *azpi* 'lower part' (< **hatz-goi*) and *bizkar* (10) 'back, hill' for **gw-* > *b-*, and *aspaldi* 'in days gone by' (< **asko-aldi* 'a long time') for **skw* > **sp*.
9. ***beilegi*** 'yellow': **behi* 'cow' + *ledi* (hyp. form of **edin* 'to become'), i.e. 'turned into the color of the cow' (cf. *berdin* 'same', *gordin* 'raw', *urdin* '*of the color of the water') > **beile.i* > *beilegi*. For *-Ø-* > *-g-*, cf. *bigun* 'weak' < Lat. *fīnu*. Although Michelena hypothesizes *hala* + *egia* > *alegia* 'as if it were', I prefer **har-le(d)i(n)* + *-a* (cf. *adina* 'as much' < **har-din* + *-na*, distributive particle), or even **-har*, as Borja Ariztimuño suggests to me. For **-r* > *-h* and **h…h…* > *Ø…h…* of *alegia*, cf. *ezker* (17) and *hibai* (22), plus *-adin*, *-aga*, *-ago* (i.e., **har* + *-din*, *-ga*, *-go(n)*).
10. ***bizkar*** 'back, hill': **goi-bar* 'upper part-lower part, inside' > **gwibar* > **bi-z-bar*. Cf. *begi* (8) for **goi-* > *b-* and *bular* 'chest, breast' (< **buru* 'head' + *bar* 'inside, lower part'), respectively; cf. *guzti* 'all' and *ditu-z-te* 'they have them' for the *-z-*.
11. ***dollor*** 'despicable'. Cf. *trollo* (Galician) in Corominas & Pascual (1980-91), under the entry 'trollo-II'. It appears that, in addition to the expected initial voicing of the stop, since *muta cum liquida* was not permitted, *-r-* and *-o-* were metathesized, and *-r* later jumped from the coda of the first syllable to the coda of the second. Similar processes are present in *leher* (36) 'to explode', *aterri* 'to clear up, to stop raining' (< Lat. *uertere*), etc. There is no change *d-* > *l-* (cf. *labur* (34) 'short'), possibly due to its late nature. Cf., however, *legen* 'ringworm', the aforementioned *leher* (36) and *leka* 'pod', as well as *lehen* 'before, first', R. *ller* 'pine tree'.
12. ***edo*** 'or (inclusive)': **e-don* 'to put', i.e. 'both X and Y', '(lit.) put X, put Y (i.e. 'let's say X, let's say Y')'. For *-n* > *-Ø* in old verbs: *ediro* 'to find' (< **eradin*, through the metathesis **ediran*), *eramo* 'to carry' (< *eraman/eroan*), *iramo* 'to throw'. Without *e-* (nor *-n*); cf. *lagun* 'companion' (< **Ø-*) / *jaun* 'Sir' (< **e-*), *lo* 'to sleep' (< **Ø-*) / *jo* 'to hit' (< **e-*), etc.
13. ***eihar*** 'dry': from Lat. *cremāre*. For *CrV* > *CVr-*, see the aforementioned processes in *dollor* (11), etc., and for *c-* > *Ø-*, *itsu* < **ketsu* < *tseku* < **caecu* 'blind', *amaña* 'shepherds' bed' (< Sp. *cabaña* 'hut'), *orein* (40) 'deer'. For the addition of *-i-*, see *saihets* (43) and those mentioned there.

14. ***etse*** 'house': **her-tz(-e)* 'to close-derived adj.(-epenthetic *-e*)'. From the same root as *hertsi* 'strict', *itxi* 'closed', *herzte/heste* 'intestine'. The *-e* comes from the reanalysis of the epenthesis of the consonant-ending stems which we can also see in *atze* 'behind', *aurre* 'before' or *Parise* 'Paris'. For the dropping of **h-* through **h...h* > *Ø...h* and ***hC*, cf. *esker* (17), *hibai* (22) (the *h-* comes from a later $*h_3 > h_1$), *ogi* 'bread' (< **hor* 'dog' + *-gi* 'thing, piece') and *otso* (41).
15. ***euskara*** 'Basque language'. According to Irigoyen, it comes from **enautsi* (which explains the *h-* of *heuskara* through $*h_2 > h_1$, after **-n-* > *-h-*). However, the root diphthong and *-tsi* are probably secondary and **e-non-tz-i* could be just as valid; cf. *inotsi* 'to speak to somebody; to rain'.
16. ***ezain*** 'ugly': Lat. *uexānu*. For *w-* > *Ø-*, cf. *aterri* (< Lat. *uertere*), *esker* 'thanks' (< Lat. *uoster*), *eskatu* 'to ask' (< Lat. *uescor*), *olde* 'thought' (< *uolontate*). For *VnV* > *VVn*, cf. *artzain* 'shepherd' (< **ardi-zani*), *hagin* (20) 'tooth', *bigun* 'weak' (< Lat. *fīnu*), etc.
17. ***ezker*** 'left': **her-tz-gu(n)*, plus *-ger* 'bad, evil'. Cf. *puzker* 'fart', *oker* 'naughty', *aker* 'billy goat', etc., with *putz* 'puff', *oin* 'foot', **han* 'animal', etc. In Lakarra (2002), I proposed simply **ez-ger*, but that is not possible, as the negation is **eze* (cf. *FHV*). For the dissimilation of *-r-*, cf. *oker* (< **or-ger*).
    For the right hand, we would have **hertz-gu(n)-hon*, with *-hon* 'good', > **hehskuon* (with *-r* > *-h*) > **ehskuon* (with Grassmann's Law) > **eskwone* > **eskone* / **eskune* > **eskoe* / *eskube* / *eskubi*. The common base is obviously the name of the hand, but not in its modern form, as believed by Tovar —that way the laminal sibilant of *ezker* would remain unexplained—, but **her-tz-gu(n)* (from **her* 'to close, to enclose', cf. *hiri* (23) 'town', + *-tz*, cf. *bor-tz* 'five', *hor-tz* 'tooth', etc.) > **hersku* > *esku*. For the group *-rzT-*, cf. *hesi* 'fence'. *Ezker*, *esku* and *eskoin* appear without *h-*, as do *etse* (14) 'house', *ogi* 'bread', *otso* (41) 'wolf', etc. Cf. *beilegi* (9) and *hibai* (22) for more cases of **-r* > *-h* and **h...h...* > *Ø...h...*
18. ***-garren*** 'Xth': **baR-en* 'inside + ines.'. For *-b-* > *-Ø-* in medial position, see *alu* 'vagina' (< Lat. *aluu*), *laur* 'four' (< *labur* (34) 'short'), *orein* (40) 'deer', etc.; for *VØV* > *VgV*, cf. *bigun* 'weak' (< Lat. *fīnu*), *hagin* (20) 'tooth', *hogen* 'fault' (< Lat. *offende(re)*), etc. On the origin and impossibility of ***batgarren* 'first', see Lakarra (2010b).
19. ***guzti*** 'all': As far as I know, no intra-Basque explanation has been proposed, and we would not sort anything out by analyzing **guz* (?) + *ti* (?). On the other hand, we have the variant B. *duzti*, which does not seem de-

rivable from *guzti*, unless we resort to an uncommon assimilation. By contrast, Lat. *tōtiu* > **tuti* > **duti* > *duzti* (cf. *bizkar* (10) 'back, hill' and *dituzte* 'they have them' for the *-z-*), with an ulterior dissimilation *d-z-t* > *g-z-t*, seems sound on both formal and semantic grounds, especially taking the peculiar appositive syntax of the particle into consideration. Alternatively, a conceivable intra-Basque explanation might be: **egu(n)* 'day' + instr. *-z* + *-ti* (adj./pros. < **din*; cf. §6.4.2.3), i.e. '*all the day' > 'all', with ulterior apheresis of *e-*.

20. ***hagin*** 'tooth': Lat. *canīnu*. For *c-* > *Ø-*, cf. *amaña* 'shepherds' bed' (< Sp. *cabaña* 'hut'), *Apodaka* (Alavese toponym, < **caput de aqua*; cf. Michelena 1987 [1981b]) or *orein* (40) 'deer'. In addition, the second shows $*h_{3/2} > h_1$-; cf. *hezur* 'bone' (< **enazur*, already proposed by Michelena), *herio* (21) 'death', *hoben* 'fault' (< Lat. *offende(re)*), *hodei* 'cloud', *heuskara* (15) 'Basque language', *haitz* 'crag, rock' (< **anetz* / **anitz*, already proposed by Michelena), etc., including *hogei* (24) 'twenty' (< **bor-gen-i*), which must have previously undergone **b-* > *Ø-* and **-r* > **-h*, and **hC* > *ØC*. The change *VnV* > *VVn* also appears in *lukainka* 'spicy pork sausage', *artzain* 'shepherd', *bigun* 'weak', etc., and *-Ø-* > *-g-* in *beilegi* (9) 'yellow', *bigun* itself, etc. For *-iu-* > *-i-*, cf. *martiu* > *marti* 'March', etc. In sum: *caninu* > **aninu* > **ãhĩhũ* > **a.ĩhũ* > **agĩhun* (with insertion of an antihiatic *-g-*) > **hagiun* > *hagin* 'tooth' and, by semantic extension, 'yew tree'.
21. ***herio*** 'death (active)': **e-lin-o*. For the root, cf. **edin* 'to (be)come' and **edun* / **elun* 'to have' (on *irudi* 'to look, to seem', cf. Lakarra 2009d). With regard to $*h_{3/2} > h_1$, see *hagin* (20) 'tooth' here; there is nothing similar (***hedario*) in **e-da-lin-o* > *jario*, in spite of belonging to the same root, nor in the rest of *e-da-ra-CVC- structures (in all of them *-h-* is at least in the initial position of the fourth syllable): hence ***hardao* 'wine', ***harrain* 'fish', etc.
22. ***hibai*** 'river': **hur* 'water' + **ban* 'to cut' + *-i*. Although *hur* has *h-* historically only in eastern areas, in Lakarra (2010b) it has been shown that it is etymological, not adventitious, as suggested in *FHV*. Thus, at some point there have been three aspirations in this compound (**huh-bahi*), in such a manner that the second dissimilated the first (**ihbahi*), as is usual (cf. *ilherri* 'cemetery', *ilhotz* 'corpse', etc.). But the anteconsonantic aspiration could not last long, and it has been the third one (**ibahi*) which has been kept after moving to the initial position, cf. *hagin* (20) 'tooth', *herio* (21) 'death'. Moreover, as Julen Manterola points out to me, in the *CSM* there is a lovely *-ibahia* (*CSM*, 1128), which renders the asterisk unnecessary.

To sum up, of the four words which belong to the same root and hence have had **h-* at some point, only *hidoi* 'marsh, swamp' and *hibai* 'river' have *h-* (**h₃ > h₁*, from **huh-dohi* and **huh-bahi*), but *idor* 'dry' (< **hur-dor*) and *ibar* 'valley' (< **hur-bar*, with both **-r > -h* and **hVhC > *ØVhC > ØVC*) do not, since in the second root of these there is no *-n-*, and hence no *-h-*. Consequently, none of the four has kept the **h-*, but the two existing *h-* are etymological, even if they have not originated in that position; cf. *harea* 'sand' (**areha* < Lat. *arēna*), *hogen* 'fault' (~ *hoben* < **obenen* < Lat. *offende(re)*), etc.

23. ***hiri*** 'town': **her-i* '*closed, surrounded' (and, by dissimilation, *hiri* > west. *huri*); cf. *etse* (14) 'house', *hesi* 'fence', *herzte/heste* 'intestine' and *hirur* 'three', *heren* 'third', for the root. The *(h)iri* 'near, right away' (historically first spatial, later temporal) kept in eastern dialects was formerly general, as shown by toponymy and old compounds: cf. *arrastirian* 'at nightfall', *goiztirian* 'at dawn', *Zufiria* (toponym) 'near the bridge', *errekairian* 'near the river', etc., just as *agiri* (> *ageri*) 'manifest, evident', from **aho-(h)iri* '*mouth-near / around' (> **a.iri > agiri*). In Romance, the Adv and the N have also the same origin (cf. Corominas & Pascual 1980-91, under the entry 'cerca', and Lakarra 2010b). Cf. also *igeri*, *uger* 'to swim', from **hur-hiri*, with the same evolution as the previous one, including the addition of the antihiatic *-g-*.
24. ***hogei*** 'twenty': **bor-gen-i* '*round-add-ptcp.' > **ohgehi* > **ogehi* > *hogei.* For *b-* > *Ø-*, see *itsaso* 'sea' (< **bits* 'foam' + *baso* 'forest'), etc. With regard to **-r* > **-h* and **hC > ØC* in *etse* (14) 'house', *ezker* (17) 'left', *hibai* (22) 'river', etc., and to *h₃ > h₁*, cf. *herio* (21) 'death', *hibai* (22) 'river', *hodei* 'cloud', etc.
25. ***ifar*** 'north': **e-da-non* 'to go' + *-(b)ar* 'inside, beneath' > **e(d)oha(n)-(b)ar > *ewha(b)ar > *ifaar > ifar*. For the metathesis *a-o > o-a*, see *joan* (33) 'to go', *unhai(n)* (45) 'oxherd', *urgatzi* (46) 'to assist', etc. For *-wh-* > *-f-*, cf. *afari* (2) 'supper' and *ifini* (26) 'to put'.
26. ***ifini*** 'to put': **edu(n)-heben-i* 'to have-here-ptcp.' > **ewhebeni* > **efe.eni* > *efini*, *ibeni*, *imini*, *ipini*, etc. For *-b-* > *-Ø-*, see *alu* 'vagina' (< Lat. *aluu*), *aulki* (7) 'chair' and those mentioned there. For the change **wh > f* noticed by Hualde (1997a), cf. *afari* (2) 'supper', *fan* (in *joan* 'to go') and *ifar* (26) 'north'.
27. ***ile*** 'hair': **enon-le*. Cf. *bele* 'raven', *erle* 'bee' for the second part (i.e. *-le*). In addition to *ule* and *elle*, we should not forget the form *ilhe*, with the changes **h₂ > h₁* and **hVR- > VRh-*: cf. *euskara* (15) 'Basque language' and *hezur* 'bone' (< **ehazur*) for the former, and *alha* 'torment, food' (cf. Rom. *hala(r)*), *anhoa* 'daily ration' (< **hanoa* < **anoha* < Lat.

annōna) (cf. *unhai(n)* (45) 'oxherd' for the details) for the latter. The *-l-*, which was fortis according to Michelena, comes from *-n + l-* (cf. *Haraluze* < *haran* 'valley' + *luze* 'long'); for *-L-* from **l + l*, cf. *bele* 'raven' < **beLe* < **bel-le.*

28. ***intsaur*** 'nut': *(p)euntze…-hur* 'meadow-hazelnut'. The first part (and its variants *Beunza*, etc.) would come from Lat. *fēnu*, as already indicated by Michelena (*FHV*, 2nd edition), and cf. *untxaur*, *eltzaur*, etc., for other variants. There are more outcomes of the same diphthong in the word families of *ile* (27) 'hair' and *iltze* 'nail'.
29. ***jabe*** 'owner': **e-da-dun-e* (**dun* 'to have'). Cf. *jaun* 'Sir', < **e-da-dun*. For the *-e*, cf. *hon* 'good' > *hobe* 'best'. Recall *utzi* 'to leave' (< **e-dutz-i*), *ilki* 'to go out' (< **e-dul-gi*) for *-d-* > *-Ø-* in the root; in the prefix, this change is general: *joan* (33) 'to go' (with metathesis), *jakin* 'to know', *jazarri* 'to attack', etc., all of them from *e-da-CVC.
30. ***jaio*** 'to be born': **e-da-din-o*. Cf. **e-ra-lin-o* (causative of *herio* (21) 'death') > **erario* > *arerio* 'enemy', **e-da-lin-o* > *jario* 'to spill' and —without prefix— **edin* > *jin* (32) 'to come'. Cf. **ardano* (3) 'wine' for the absence of $*h_4 > h_1$.
31. ***jarein*** 'to spill': **e-da-len-i*, not **jar-egin* as believed by Michelena (1987 [1977]), conceiving a first "dark" or "unknown nominal element" (in his own words [translation by the editor, MMA], although he should rather have said "nonexistent"). Recall that, in autochthonous words, *j-* comes from the verbal prefix **e-* and that *jarein* —as envisaged by Michelena himself for *jardun (inardun)* 'to work on, to devote oneself to' and *jardetsi/inardetsi* 'to discuss'— should have also yielded ***jargin*, without the pre-radical ***e-*. For the same reason, it cannot be accepted that *jaramon* 'to pay attention' comes from **jar-emon* (nor from *chera emon*, recall *etxaramon* 'to ignore, to take no notice of', from ez *jaramon* 'no attention'). Instead, we have to start from a verbal CVC with more structure to the left of the root. For other derivatives of the same stem, cf. *urretx* 'hazelnut' (< **huR-len-i-tz* '*hazel tree-planted') and *hareitz* 'oak tree' (< **e-len-i-tz*) and *leizar* 'ash tree' (< **len-i-tz-har*).
32. ***jin*** 'to come': **e-din*; cf. *jaio* (30) 'to be born', *jario* 'to spill', *herio* (21) 'death', etc., and *dadin*, *zedin*, etc. (synthetic forms of **edin*). In Romance (and English), 'to come' and 'to become' have also the same root; cf. RS. (263) *Lotsaga nindin*, *ogi-z ase nindin* 'I became shameless and got sated with bread', etc., and Lakarra (2006c).
33. ***joan*** 'to go': **e-da-non*; cf. the variant *johan* and especially finite forms like *doha* '(s)he goes', *zihoan* '(s)he went', etc., where the aspiration is more striking due to their lack of accent. The same metathesis is present

in *unhain* (45) 'oxherd' and other derivatives of the same root: *erho* / *eran* 'to kill' < **e-ra-non* (> **erahon* → *hara-gi* 'meat') and *eroan* / *eraman* 'to carry' (**e-da-ra-non* > **eranoan* > **eraoan*, without *-h-* > *h-*, because **-n-* is in the fourth syllable; cf. **ardano* (3) 'wine'). The variant *fan* comes from **edohan* > **ewhan* (cf. *afari* (2) 'supper', *farre* 'laugh', *ifar* (25) 'north', *ifini* (26) 'to put') and later dropping of **e-*.

34. ***labur*** 'short': **da-bur*. The prefix *la-* is also in *labain* 'slippery', *labar* 'edge of a precipice', *lagun* 'companion', etc., and *-bur* in *samur* 'tender', *apur* 'a little', *zimur* 'wrinkle', and perhaps *urri* 'scarce, october' < **e-burr-i. l-* < **d-* also occurs in *lats* (35) 'cascade, rapids', *lehen* 'first, before', etc. Apparently, *laur* 'four' is an evolution of *labur*; cf. *alu* 'vagina' (< Lat. *aluu*), *bular* 'chest, breast' (< **buru-bar*), with *-b-* > *-Ø-*. As was explained in (2010b), in several Amerindian languages counting on one's fingers begins neither with the little finger nor with the thumb (the two at the extremes), but with the forefinger, and hence the little finger turns out to be neither the first nor the last, but the fourth one (cf. Epps 2006).
35. ***lats*** 'cascade, rapids': < **dats*; cf. *aldats* 'slope' (< *alde* 'side' + **dats*), *adats* 'hair' (< **da-dats*, with reduplication; cf. *adar* (1) 'horn', *odol* (39) 'blood'), *laster* (*ur* __) 'rapids', etc., and also **d-* > *l-* in *labur* (34) 'short' and *leher* (36) 'to explode'.
36. ***leher*** 'to explode': Lat. *tremere* > **terme* > **lerme* > **lerne* > **lener* > *leher*. We have *t-* > *d-* > *l-* in *leka* 'pod' (< Lat. *thēca*), *letagin* 'canine' (< **detagin* < **betagin*), and *CrV* > *CVr* in *dollor* (11) 'despicable' and *izter* 'thigh' (< Sp. *ristra* 'string (of garlic/onions)', or even —and perhaps semantically closer— Sp. *ristre* 'iron sheath for the spear'), *bazter* 'corner' (< Lat. *praesēpe* 'manger'). We have also a jump of *-r* from the coda of the first syllable to the coda of the second in *aterri* 'to clear up' and *dollor* (11) 'despicable'.
37. ***mahai(n)*** 'table': **goi* 'upper part' + **da-don-i* (> **dodani*) 'put (ptcp.)' > **gwi-(d)odani* > **bionani* > **bonani* > **banani* > **bãhãhĩ* > *mahain*. For **gw-* > *b-*, see *bizkar* (10) 'back, hill'; for the second part, cf. *unhai(n)* (45) 'oxherd'.
38. ***nagusi*** 'main, oldest': **da-dun-tz-i*; cf. *jauntzi* 'to get dressed' < **e-da-dun-tz-i*. For a similar pair with and without **e-*, cf. *laino* 'fog, cloud' (< **da-din-o*): *jaio* (30) 'to be born' (< **e-da-din-o*). Other non-Proto-Basque *n-* are those in *neska* 'girl', *nigar* 'crying', *narru* 'skin'.
39. ***odol*** 'blood': **dol* → **do-dol* (with reduplication; cf. *adar* (1) 'horn, branch', *adats* 'beautiful', *eder* 'beautiful', etc.), with identical process

**C-C > Ø-C*; also *ahal* ‘can, to be able to’, *ihintz* ‘dew’, *ohol* ‘plank’, etc. Other members of the same family are *erori* ‘to fall’ (< **e-ra-dol-i*), *jori* ‘abundant’ (< **e-dol-i*, which has a different origin from that of *heuragi/ugari* ‘abundant’ < ‘*too much’, < **e-ra-dun-egi*; cf. *FHV*), *oskol* (in B, and hence it might come from *ozkol*) ‘shell’ (< **zotz* ‘twig’ + *-dol*).

40. ***orein*** ‘deer’: **cer(u)u-oin* ‘deer-handle’ > **er(u)-oin* > **or-ein* > *orein*. We find an identical metathesis in *hodei* ‘cloud’. For *-u* > *-Ø-* see *bular* ‘chest, breast’ (< **bur(u)* ‘head’ + *bar* ‘inside, beneath’), *burhezur* ‘cranium’ (< **bur(u)* ‘head’ + *hezur* ‘bone’), and for *-b-* > *-Ø-* see *alu* ‘vagina’ (< Lat. *aluu*), *labur* (> *laur*) ‘short’, *-garren* (18) ‘Xth’. For the second element (*-oin*), recall *kirten* ‘handle’, etc. (*FHV*: 508-509).
41. ***otso*** ‘wolf’: **hortz-so* ‘big eyetooth’; cf. *amaso* ‘grandmother’, *aitaso* ‘grandfather’, as well as *baso* ‘forest’ (< **bar-so*) and its derivatives, as well as *hortz* ‘eyetooth’ (< **hor* ‘dog’ + *-tz* ‘-ish’, i.e. a derivational suffix). For the consonant group (*rtzC* > *ts*), cf. *etse* (14) ‘house’, *esku* ‘hand’ and *ezker* (17) ‘left’ (with dissimilation of rhotics in the latter); we have **h-* > *Ø-* in all of them; cf. *beilegi* (9) and *hibai* (22) here.
42. ***sabel*** ‘abdomen, stomach’: *sa-bel* ‘inner-black’. For *sa-*, cf. *sabai* ‘loft’ (< **sa-ban-i* ‘*cut (ptcp.) inside / underneath’), *sa(i)hets* (43) ‘rib, side’, *sakon* ‘deep’, *samin* ‘severe pain’, *samur* ‘tender’. For *-bel*, cf. *harbel* ‘slate’, *horbel* ‘fallen leaves’, *ubel* ‘bruise’ (< **uh-* < **hur-*), *gibel* ‘liver, behind’, etc.
43. ***saihets*** ‘rib, side’: **sa-hets* ‘inner-closed’. For the prefix, cf. *sabel* (42) ‘abdomen, stomach’, and for the root *hertze* ‘intestine’, *hertsi* ‘closed, strict’, *etse* (14) ‘house’, etc. With regard to the addition of *-i-*, cf. *eiho* ‘to grind’, *eihar* (13) ‘dry’, *aihotz* ‘sickle’, etc.; cf. Lakarra (2009c). To these, we can add *oihu* ‘cry’, from Gasc. *hou-hou* ‘interj., cri comme oh! holà! pour hêler ou chasser’ (cf. Palay 1991, under the entry ‘hou-hou’), and probably *deihadar* (cf. *atzamar* (6) ‘finger’).
44. ***sei*** ‘six’. The proposals which assume a borrowing process from Romance or Iberian are unfounded, as Lat./Rom. (or Iberian) *s-* would have yielded Basq. *z-* in words with a single sibilant. Moreover, the diphthong *ei*, which appears in all periods and dialects, would be unexpected if it were a PB. word or an old loan word. In fact, an old *ei* should have yielded ***si* in all or almost all dialects (cf. *liteke* ‘(s)he could be’ < *leiteke*), but such a form is unattested. The cases of a kept *ei* are clearly secondary, coming from loan words, old hiatuses, old bisyllabic combinations in which an intervocalic voiced stop has dropped out, depalatalization processes, vocalizations of a *-k-*, etc. (*FHV*: 104-106).

Instead (cf. Lakarra 2010b), I propose that Basq. *sei* 'six' comes from **seni* 'son, child, boy' (which actually would not require any asterisk, as we have Aquit. *SENI-*), later > *se(h)i* (with the western variants *segi, seẽ, seiñ*). A semantic derivation '*son, child' > 'six' is paralleled in Amazonian languages (cf. Epps 2006), where the nouns for all numbers from 4 to 10 contain an indication of the existence (4, 6, 8, 10) or not (5, 7, 9) of 'brother', since, when counting on one's fingers, even numbers are a counterpart (lit. a 'brother') to the previous odd one. Formally, the absence of aspiration in *sei* '6' (as opposed to *se(h)i* 'child') even in those dialects which maintain this feature must be due to the clitic nature of numerals. In all other respects, this proposal —unlike the hypothesis of borrowing— satisfactorily accounts for the resulting comm. *sei*.

45. ***unhai(n)*** 'oxherd': **e-da-don-i* (**don* 'to put ~ to take') > **edonani* (with vowel metathesis, cf. *joan* (33) 'to go' and *urgatzi* (46) 'to assist', and assimilation *-doni* > *-noni*) > **edõhãni* > **euhani* > **hunai(n)* (with metathesis **h_3 > h_1*) > *unhain* (with metathesis *hVR- > VRh-*) and *ul(h)ain* (with dissimilation). Cf. *urdain* 'swineherd' (< **urde* 'pig, hog' + **edodani*) for the final and *urgatzi* (46) 'to assist', *ulertu* 'to understand' / *igerri* 'to guess' for the *u-* (cf. *zauri* (47) 'wound'). In *mahai(n)* (37) 'table' we have another derivative of the same root.
46. ***urgatzi*** 'to assist': **e-da-ra-gotz-i*; cf. *joan* (33) 'to go' and *unhai(n)* (45) 'oxherd', for the metathesis *a-o > o-a*, and *arrain* (4) 'fish', *arreba* (5) 'sister of a man', for the *-R-*.
47. ***zauri*** 'wound': Lat. *sanguine* > **zaungine* > **zaugihe* > **zaugii* > *zauri*. Cf. *sorotsi* 'to assist, to help' (< **so-egotzi* 'look-to throw'), and also *ulertu* 'to understand' (< **urertu* < **ugertu* (cf. *igerri* 'to guess'), both from **edun-berri* 'to have news'), for *-g- > -r-*. We have the same dropping of a nasal before a velar in *iguzu / indazu* (synthetic forms of **nin* 'to give'), *ukitu* < *unkitu* 'to touch', etc.
48. ***zemai*** 'threat': Old Cast. *menaza* > **mezana* > **zemana* > **zemaa* > **zemae* > *zemai*. Basque has also *beaza* and *mehatxu*, of identical origin, but they are not attested in the same varieties; *amenazu, amenazatu*, etc. —with late Romance agglutination of *a-*—, by contrast, can appear with —i.e. in the same variety as— either of them.
49. ***zezen*** 'bull': **zen* → *ze-zen* (with reduplication; cf. *adar* (1) 'horn, branch', < **da-dar*, and *zozo* 'thrush', among many others); cf. *zekor* 'young bull' < **zen-gor*.
50. ***zortzi*** 'eight': from *zorrotzi* 'sharp' (attested in eastern varieties with the meaning of '8', cf. *OEH*). The group comes from **zorrortzi* (< **zur-hortz-i* 'wood-tooth-ptcp.'), since, were it old, then *-s(t)-* would be ex-

pected in eastern dialects and Souletin, as well as (some varieties of) Roncalese (cf. *beste / bertze* 'other'). As was pointed out in Lakarra (2010b), there is a relationship between the words for '8' and 'sharp' in many other languages (beginning with the Indo-European family), due to the fact that, when counting on one's fingers, the eight knuckles of both hands, employed to represent that number, take on the appearance of two mountain ranges.

6. Noun Morphology

Ekaitz Santazilia
(Public University of Navarre, UPNA/NUP
/ University of the Basque Country, UPV/EHU)

6.1. Introduction*

This chapter deals with the functioning and grammaticalization of the Basque case system, including the marking of determination and number in the noun phrase. First, I present a synchronic description of that system, and later I analyze the data from a diachronic perspective, summarizing the different hypotheses proposed by Bascologists, adding to these the latest contributions, and giving an account of the most recent avenues for research.

6.2. Description of today's Basque noun declension system

The Basque noun phrase presents a canonical structure which can be linearly represented in the way shown in Figure 6.1.[1]

Det.1	Noun	Adjective	Det.2	Number	Case[2]

Figure 6.1. Canonical structure of the Basque NP

As can be observed in (1a), it is the whole noun phrase that is marked for number and case in Basque, the morpheme being attached to the end of the

* I wish to thank Maider Bedaxagar, Iván Igartua, Joseba Lakarra, Mikel Martínez-Areta and Patxi Salaberri-Zaratiegi for their interesting comments on previous versions of this work, and Mikel Martínez-Areta for translating the text. The work has been carried out thanks to the Grants for the Training of Researchers funded by the Basque Government (BFI-2010-221), to the Basque Government's consolidated research group "Historia de la Lengua Vasca y Lingüística Histórico-Comparada" (HLMV-LHC), Ref. GIC 10/83, IT 486-10, and to the Formation and Research Unit LingTeDi (UFI11/14) of the UPV/EHU.

1 In this section, I am following Lafon (1999 [1970]), Trask (*HB*) and Martínez-Areta (2009a).

2 From the point of view of most morphologists and syntactic theoreticians, this formula would not really be linear, nor would all of its elements be interconnected in the same manner. For the sake of clarity, however, let me be allowed to resort to this simplification.

whole NP, unlike in languages like Latin (1b), which mark each element of the phrase:

(1) a. *Bi* *gizon* *handi* *hori-ei*
two man big that-DAT.PL
'To those two big men'

b. *Ist-is* *duo-bus* *magn-is* *homin-ibus*
that-DAT.PL two-DAT.PL big-DAT.PL man-DAT.PL
'To those two big men'

Basque has a single declensional paradigm, which is applied to all nouns (and adjectives) alike, even if the union between noun and declensional morpheme triggers some phonological changes. Thus, a vowel-ending noun presents the paradigm shown in Table 6.1 in Standard Basque.[3]

Table 6.1. Noun declension of vowel-ending nouns

etxe 'house'	Definite		Indefinite
	Sg.	Pl.	--
Absolutive	*etxe-a-Ø*	*etxe-ak-Ø*	*etxe-Ø*
Ergative	*etxe-a-k*	*etxe-ek*	*etxe-k*
Genitive	*etxe-a-r-en*	*etxe-en*	*etxe-r-en*
Dative	*etxe-a-r-i*	*etxe-ei*	*etxe-r-i*
Instrumental	*etxe-a-z*	*etxe-ez*	*etxe-z*
Inessive	*etxe-an*	*etxe-eta-n*	*etxe-ta-n*
Allative	*etxe-ra*	*etxe-eta-ra*	*etxe-ta-ra*
Ablative	*etxe-tik*	*etxe-eta-tik*	*etxe-ta-tik*

Although not in an entirely precise manner, by and large definite forms are used whenever we refer to concrete nouns which are known to both the speaker and the addressee, whereas the indefinite paradigm is reserved for elements which are not determined or not concrete:

3 There exist dialectal differences in the Basque declension, but as Lafon (1999 [1970]: 167) recalls: "Dès les plus anciens textes, c'est-à-dire dès XVIe siècle, elle est la même, ou à peu près la même, qu'aujourd'hui", with notable exceptions, such as the soc. *-gaz* and the abl. *-rean* in the west, or the dat. pl. *-er* in the east (Michelena 1987 [1981a]: 301).

(2) a. *Bi* *etxe-Ø* *eros-i* *d-it-u-t*
two house-ABS.INDEF buy-PFV PRS-ABS3PL-ROOT-ERG1SG
'I have bought two houses'

b. *Bi* *etxe-ak* *eros-i* *d-it-u-t*
two house-ABS.PL.DEF buy-PFV PRS-ABS3PL-ROOT-ERG1SG
'I have bought the two houses'

As can be observed, only the definite paradigm distinguishes number (sg. or pl.). In (2b), it is evident that when we refer to more than one house (two, in this case) and they are definite, the verbal marker must be pl. The same holds true for most dialects and Standard Basque when they are indefinite, as in (2a), but in far eastern varieties and some old western texts, the verbal marker would be sg. in such a case.

In the paradigms of this section, I have provisionally included a total of eight cases. As a matter of fact, we will see that Bascologists disagree as to the exact number of cases in Basque, either because some of them consist of several suffixes (e.g. the destinative *etxe-arentzat* 'for the house' can be formally split into the genitive *-aren* plus the suffix *-tzat*, which is essive/translative if it is directly attached to the root), or because some cases do not seem to behave like the others.[4] The cases not included in these paradigms have been discussed in §6.5 and §6.6. Be it as it may, it has been stated that Basque has a relatively rich case-system in comparison with many other languages of the world (cf. Iggesen 2011), but it is also true that, as e.g. in the Uralic family, a considerable number of those cases are derived by attaching postpositions to other (primary) cases (cf. §6.5).

As I have said, Basque has only one declensional paradigm, unlike other languages like Latin. This has sparked discussions among Bascologists about the existence of a declension proper (Rotaetxe 2008: 591). If we consider that we have one single paradigm, and that this is formed, at least in origin, by means of postponed agglutinated morphemes, it might be more suitable to speak about postpositions, and reserve the term *declension* for inflecting languages. Here, however, we cannot get tangled up in theoretical issues, and for the sake of convenience I will employ the term declension in the rest of the paper.

Consonant-final nouns are declined in the way shown in Table 6.2. It should be pointed out that, in that table, the pl. and indef. paradigms are similar at the phonological level (except for the abs. and dat.), due to the epenthesis of the eu-

4 For a description of the meaning of these and other cases, cf. Trask (*HB*: 92 ff.).

phonic *-e-* which is inserted in the indefinite paradigm in order to prevent the clash of two consonants.

Table 6.2. Noun declension of consonant-ending nouns

jolas 'game'	Def.		Indef.
	Sg.	Pl.	--
Abs.	*jolas-a-Ø*	*jolas-ak-Ø*	*jolas-Ø*
Erg.	*jolas-a-k*	*jolas-ek*	*jolas-e-k*
Gen.	*jolas-a-r-en*	*jolas-en*	*jolas-en*
Dat.	*jolas-a-r-i*	*jolas-ei*	*jolas-i*
Instr.	*jolas-a-z*	*jolas-ez*	*jolas-e-z*
Ines.	*jolas-e-an*	*jolas-eta-n*	*jolas-e-ta-n*
All.	*jolas-e-ra*	*jolas-eta-ra*	*jolas-e-ta-ra*
Abl.	*jolas-e-tik*	*jolas-eta-tik*	*jolas-e-ta-tik*

The cases from the abs. to the instr. present *-a(r)-* in their singular paradigm, which is simply the article, and the rest of the cases have *-(e)ta* in the pl. and indef. The latter are the so-called 'local' cases, which —as will be shown in §6.3.2— display special features.

Finally, in Table 6.3 I present the paradigm of proper nouns, which are always definite.

Table 6.3. Noun declension of proper nouns

	Def. Sg.	
	Vowel-ending	Consonant-ending
Abs.	*Iruri-Ø*	*Gasteiz-Ø*
Erg.	*Iruri-k*	*Gasteiz-e-k*
Gen.	*Iruri-r-en*	*Gasteiz-en*
Dat.	*Iruri-r-i*	*Gasteiz-i*
Instr.	*Iruri-z*	*Gasteiz-e-z*
Ines.	*Iruri-n*	*Gasteiz-e-n*
All.	*Iruri-ra*	*Gasteiz-(e-r)a*
Abl.	*Iruri-tik*	*Gasteiz-(e)-tik*

Proper nouns lack articles, which on the contrary are neatly present in the singular paradigm of common nouns: cf. abs. *lagun-a-Ø* vs. *Iruri-Ø*. On the other hand, the suffixes beginning with a vowel are directly attached to consonant-ending roots (*Gasteiz-i*), but if the root ends with a vowel, an epenthetic *-r-* shows up which we also saw in the paradigm of common nouns (cf. *Iruri-r-i*, and also *etxe-r-i*, *etxe-a-r-en*, etc.).

Apart from the suffix *-(e)ta-*, one of the peculiar features of the local cases is the morphological distinction between animates and inanimates. If we bring this distinction into Tables 6.1 and 6.2, a revision of the two would be Table 6.4.

Table 6.4. Animate and inanimate paradigm of local cases

	Sg.		Pl.		Indef.	
	Inanimate	Animate	Inanimate	Animate	Inanimate	Animate
Ines.	*itsaso-an*	*lagun-a-(ren)-ga(-)n*	*itsaso-eta-n*	*lagun-en-ga(-)n*	*itsaso-ta-n*	*lagun-(en)-ga(-)n*
All.	*itsaso-ra*	*lagun-a-(ren)-gan-a*	*itsaso-eta-ra*	*lagun-en-gan-a*	*itsaso-ta-ra*	*lagun-(en)-gan-a*
Abl.	*itsaso-tik*	*lagun-a-(ren)-gan-dik*	*itsaso-eta-tik*	*lagun-en-gan-dik*	*itsaso-ta-tik*	*lagun-(en)-gan-dik*

The animate paradigm is built upon the genitive, to which a morpheme is attached, which nowadays appears as *-ga(n)*. In modern Standard Basque, this genitive is optional in the sg. and indef., but obligatory in the pl., probably to distinguish it from the indef.[5] In the continental varieties, instead of these forms we find the postposition —of uncertain origin— *bait(h)a*,[6] which takes the genitive (*EMH*: 114), although in old texts it also appears with the abs. Thus, the inessive sg. animate would be *lagun-aren baithan* 'in the friend', the abl. pl. *lagun-en baitha-tik* (or *baitha-rik*, cf. §6.4.2.3) 'from the friends', etc. We also come across this postposition with the meaning of 'house' in the same geographical domain, but Trask (*HB*: 208) considers this to be a specialization of what must have had a generalized local use.

6.3. The article, definiteness and number in noun declension

It is commonly accepted that the article, placed between the root and the case marker, is of recent introduction in the language (Schuchardt 1947 [1923]: 46, Michelena 1987 [1978], Manterola 2009a). It shows differences of use between dialects, the eastern ones being the most reluctant as to its use. I agree with

5 Nonetheless, these plural forms of animate nouns formed by *-ga(n)* upon the genitive are relatively recent (cf. Castaños-Garay 1979: 12 ff.). They first appear in Old Labourdin and Ber. (17th century), and seem to spread centrifugally. In 16th century texts, we always find animate nouns declined as inanimates, as in RS. *gizon-etan* 'in the men', Laz. *andra-etan* 'in the woman', Dch. *gizon-etarik* 'from the men', etc.

6 It might come from *bait da* 'which is' (cf. Creissels & Mounole 2011: 179).

Lafon (1999 [1954a]: 216) that the heterogeneity of number marking in Basque (depending on the root being vowel- or consonant-ending), and the differences in the formation of the indef., sg. and pl. paradigms, denote a progressive grammaticalization, which makes use of different elements.

In today's Basque, the article not only marks the definite / indefinite division, but also bears the feature of number (Lafon 1999 [1954a]: 211).[7] Let us take as an example the grammatical cases (absolutive, ergative, dative, instrumental, and genitive),[8] which are the ones that most clearly allow us to detect the presence of the article, at least in the sg. We shall then analyze the morphology of local cases, which presents certain peculiar features.

6.3.1. The grammatical cases

Due to its lack of overt case marker, the abs. case (*-Ø* morph in the indef.) is the clearest for discussing the issue of the article. Hence, I shall take it as a starting point. As can be seen in Table 6.1, we have abs. *etxe* for the indef., *etxe-a* for the def. sg., and *etxe-ak* for the def. pl.

The most widely accepted view, probably since Azkue's *MV*, is that the article —at least the one which grammaticalized from the 3rd grade demonstrative— came about in Basque at the same time as in the Romance languages, and in a parallel manner: from demonstratives (Lafon 1999 [1970]: 168, Michelena 1987 [1978], Trask *HB*: 199). According to the data put forward by Manterola (2006, 2008), we must locate that grammaticalization to around the 10th century AD. Thus, the *-a* of the sg. article would simply be the grammaticalized form of the 3rd grade demonstrative **har*.

There are several arguments in favor of this theory. To begin with, in the Aquitanian inscriptions (from approximately the 1st-3rd centuries AD) there is no trace of the article.[9] Secondly, in western dialects the 3rd grade demonstrative is still *a*: *etxe a* 'that house there'. Thirdly, Manterola (2008) has brought into the discussion several late medieval place names from Alava in which the article,

7 Lafon himself (ibid.: 217) suggests, not without hesitation, that an indefinite plural may also have existed, but I believe that he bases this on a wish to obtain a symmetric paradigm rather than on linguistic evidence.

8 Although there are several possibilities for grouping and naming the Basque cases, here they will be grouped into grammatical ones and local ones (pseudo-declensional cases are left out of either group, cf. §6.6). Within the grammatical cases, I shall call those which trigger an agreement marker on the verb "core cases" (i.e. abs., erg. and dat.).

9 This argument, however, is not conclusive, as the attested data in those inscriptions are proper nouns, which would not be expected to take the article, even if this had already grammaticalized.

not fully grammaticalized in its modern form, keeps the aspiration, as in *Elhorzaha*, *Artazaha* (today's *Elorza*, *Artaza*), etc.[10] Fourthly, there is evidence that in a not very remote period, the 3rd grade demonstrative was not the only one grammaticalized as an article: we have medieval names like *Lope Ederro* (instead of *ederr-a* 'handsome'), or present-day family names like *Ezquerro* which may reasonably be derived from an article of proximity coming from the 1st grade demonstrative **hau(r) > ho(r) > -o* (Irigoyen 1981: 385-9, §7.2.4).[11] In western varieties, there is still an article in *-ok*, side by side with the common *-ak* (*HB*: 94).[12] From the 2nd grade demonstrative *(h)ori*, we would derive forms like *gorpuzori* (*gorputz* 'body'), found at a *fase sparita* in Dechepare's LN. dialect (16th century), and in somewhat later Labourdin authors (Michelena 1987 [1979b]: 450 ff., Castaños-Garay 1979, *HB*: 199).

As for the lexical origin of this **har*, Lakarra (2008a: 484) has proposed an interesting theory. After showing the relevance of the grammaticalization of serial verbs in the formation of numerous affixes in today's Basque, and noticing the origin of the article from a lexeme close to the verb 'take' in a number of languages, he has proposed that lexematic origin for Basque **har* 'that there'. As a matter of fact, this form would not require any asterisk, as the verb *har* 'to take' still presents that form today.

The forms without article have lost ground over history (Lafon 1999 [1954a]: 212), but, as I have said, eastern varieties are more conservative as far as the use of the article is concerned. As we move westwards, the presence of the article is found in increasingly more contexts. On the other hand, the penetration of the article has not been homogeneous, and has not run parallel to the Romance languages, for the Basque article is nowadays used in contexts where other Romance languages do not require it (Manterola 2012).

Without going into details (cf. *EMH*), I shall simply mention that the article penetrates more easily into the subject complements —or predicate adjectives—

10 It may be mentioned, in passing, that the presence of this aspiration renders any attempt to relate the Basque article to the Berber one *-a* impossible. Cf. Berber *argaz-a* 'the man' (Núñez-Astrain 2003: 288). For more problems concerning this and other genetic relationships, cf. Lakarra (2006b).

11 Although in my opinion it is not as strong as Irigoyen's proposal, the possibility —already hinted at by Michelena (1987 [1969]: 138)— that *-o* has arisen by analogy with the gender distinction in Romance, cannot be ruled out. Thus *ederra* was interpreted as feminine and *ederro* as masculine, especially in already Romanized areas. The feminine nouns that show up in both Irigoyen's and Michelena's works are very few, and among them, those which present that form in *-o* are practically nonexistent.

12 And hence corresponding case forms across the paradigm such as ines. *etxeotan*, soc. *etxeokin*, etc.

at *individual level* (Zabala *GB*: 329)[13] and into grammatical core cases (especially erg. or dat.),[14] and more easily into countable nouns than into uncountable ones in object function (Manterola 2006, Santazilia 2010), or even in word-listing contexts. In this respect, two attestations worth comparing are Aimery Picaud's word list and that of Arnold von Harff (cf. *TAV*: 49-51, 63-65, respectively, and §3.3), as both were gathered in Navarre —the former in the 12th century and the latter in the 15th century— and appear to reflect the spread of the article in progress.[15] As Table 6.5 illustrates, whereas in Picaud's vocabulary countable concepts are given in the def. sg. form and uncountable concepts are given in the indef. form (with the exception of *ur-ic*, in the partitive case), in von Harff's list uncountable concepts also bear the article, with the exception of *aragui*. Note that in some cases the listed words are exactly the same.

Table 6.5. Aimery Picaud's and Arnold von Harff's word lists

	Picaud (12th c.)	Von Harff (15th c.)	Translation
countable nouns	*Urci-**a***		'sky (/ God)'
	*belaterr-**a***		'presbyter'
	*eregui-**a***		'king'
	*andre-**a***		'Lady'
	*iaon-**a***	*(hytzoko)sanm-**a***	'Sir'
uncountable nouns	*araign-**Ø***		'fish'
	*o(r)gui-**Ø***	*oge-**a***	'bread'
	*aragui-**Ø***	*aragui-**Ø***	'meat'
	*ardum-**Ø***	*arduw-**a***	'wine'
	*ur-**ic***	*oyr-**a***	'water'
	*gari-**Ø***		'wheat'
		*oluw-**a***	'oats'
		*gaz-**a***	'salt'
[*TAV*: 49-51, 63-65; adapted version, not including all words]			

13 In fact, at *stage level* the article still does not appear today: *irakasle-Ø dago* 's/he is as a teacher' vs. *irakasle-a da* 'he is a teacher'.

14 However, in the historical period we still have cases like Ax. (L, 17th century) erg. indef. *aingiru-k* (*aingiru* 'angel' + erg.), which would be ungrammatical today, unless we use them with an indefinite determiner (*hainbat aingeru-k* 'some angels').

15 Another word list often mentioned within this set of attestations is that gathered by Lucius Marineus Siculus in the 16th century (*TAV*: 146-147 and §3.3). In it, all items are given with the article. I have left it out of Table 6.5 because it corresponds to the western dialect, where the spread of the article took place faster, and hence it is not exactly comparable to the testimonies of Picaud and von Harff.

This earlier emergence of the article in countable nouns is entirely expected on typological grounds. This is no doubt due to the fact that the notion of singularity or plurality is semantically easier to capture in a countable element than in an uncountable one (Martínez-Areta 2009a: 76). It should not be forgotten that in Basque, number is morphologically expressed only by the article, a feature inherited from the demonstrative from which it comes. Hence, the only way to emphasize plurality is to use it, thus giving rise to an ambiguity: a form like *arraultze-ak*, with the plural article, can mean both 'eggs' and 'the eggs'. There is no overt way of expressing indefinite plurality. Hence, as the marking of number has become more important than that of definiteness, the article has even crept into contexts which are semantically indefinite, in order to mark singularity or plurality (Martínez-Areta 2009b: 254), as this example of *arraultzeak* shows.

Example (3), taken from a corpus from the end of the 19th century (Santazilia 2010: 192 ff.), shows a synchronic alternation. In (3a), uncountable nouns like 'bread' and 'wine' are expressed by the indefinite form, but so are those of plural semantics like 'eggs', undoubtedly due to their indefinite nature. It is in (3b) that we find a stage at which the morphological marking of plural is present in semantically plural phrases like 'eggs'. As already mentioned, definiteness and number go hand in hand, but in (3b) the former is obviated by the growing importance of the latter, thus marking in the same way as a definite NP a noun which is semantically indefinite. At the following stage, uncountable nouns will also end up being definite:[16]

(3) a. *Betea da mutil-ez zeinek baitute ogi-Ø eta*
full is boy-INSTR.PL who have bread-ABS.INDEF and

arno-Ø, arraultze-Ø eta gasna
wine-ABS.INDEF egg-ABS.INDEF and cheese.ABS.INDEF(?)

b. *Betea da mutil-ez zeinek baitute ogi-Ø eta*
full is boy-INSTR.PL who have bread-ABS.INDEF and

arno-Ø, arraultze-ak eta gasna
wine-ABS.INDEF egg-ABS.PL and cheese.ABS.INDEF
'It is full of boys who have bread, wine, eggs and cheese'

16 For a typological overview of all this, cf. Martínez-Areta (2009b: 254 ff.).

Therefore, it is evident that, before the creation of the article, neither number nor determination was expressed by agglutinated suffixes. In fact, this had already been proposed by Gavel (1929: II, 44).[17]

It is only logical, then, to consider that at first the article arose to create an opposition between the indefinite and the definite, not to mark number (*HB*: 199), although the latter function is inherent in the article, thanks to its former nature as sg. demonstrative. In the noun paradigm, we would initially only have markers like the *-Ø* of *etxe-Ø* in the abs. indef., and the *-a* of *etxe-a* in the def., the latter with a connotation of singularity inherited from the demonstrative. If we look at Tables 6.1 and 6.2 in their sg. paradigm, we can see that it is precisely in these grammatical cases[18] that the article *-a(r)-* is perfectly identifiable.

What about the pl. article *-ak*? It is commonly accepted that this —and actually the whole pl. paradigm—[19] arose later than the sg. article.[20] Michelena (1987 [1971b]: 145) suggested —although not very forcefully— deriving *-ak* from *-aga*,[21] a suffix present in place names (*Iturri-aga*, *Elorri-aga*, etc.). The

17 The question mark in (3a) requires an explanation. On phonetic grounds, we cannot know for sure whether in *gasna* the sg. article (*-a*) is present or not, for it is an *a*-ending root. However, its coordination with *arraultze-ak* suggests that it has the article (since in Basque, two coordinated NPs can take the article, and so can only the second, but it would be ungrammatical if the first had it and the second did not). In some varieties, the indef. form and the def. sg. one are homophonous in such roots. In others, by contrast, there are some phonetic differences between the indef. and the def. form. In the west and a part of the center, a dissimilation *-a* + *-a* > *-ea* occurs when the article *-a* is attached to the final *-a* (in the whole def. sg. declension, except in the ines.), as in *alaba-a* > *alabea* 'the daughter' ~ *alaba* 'daughter'. In Souletin, the *-á* of the definite abs. sg. is accented, unlike that of the indef. (*alhabá* ~ *alhába*, respectively), and in Salazarese, an epenthetic *-r-* is inserted between the final *-a* and the article (*alaba-r-a* ~ *alaba*, respectively).

18 Perhaps with the exception of the genitive, these cases can be considered the oldest ones (*HB*: 201).

19 In fact, this shows a greater dialectal diversity than the singular paradigm (cf. Castaños-Garay 1979: 65).

20 Michelena (1987 [1981a]: 48-9) puts forward an internal argument in favor of this. In western varieties, *a*-ending roots undergo a dissimilation *-a-a* > *-ea* when taking the sg. article: *hanka* 'leg' + *-a* > *hankea* 'the leg'. This does not occur with the pl. article, where we find *hankak* (or *hankaak* in Old B and Laz.), not ***hankeak*, a fact explained by Michelena by suggesting that when the pl. article was grammaticalized the dissimilation was no longer operative.

21 Earlier authors like Albert León and Henri Gavel (1920: 339, 362; 1929 II, 53-55) had already envisaged this. As will be shown below, today we know that the bisyllabic nature of this morpheme renders its existence in Proto-Basque impossible, and requires segmentation.

evolution would be as follows: *-aga* > **-ag* > *-ak*, i.e. first apocope of *-a*, and then devoicing of the stop in final position.[22]

Nevertheless, as de Rijk (1998 [1981]: 218) recalls, Michelena himself poses problems to this explanation. On one hand, this suffix is more locative than abundantial,[23] and hence it presents semantic problems to being linked to the notion of plurality. On the other hand, as already mentioned, in western varieties the dissimilation *-a*-ending root + *-a* > *-ea* does not occur in the plural, but it does occur if the attached element is the suffix *-aga*, as we have the family and place name *Oleaga*, from *ola* 'factory, forge' + *-aga*.[24]

Let us recall that we start from a linguistic situation in which neither determination (definite / indefinite) nor, consequently, number is marked by means of clitics (Martínez-Areta 2009b: 250). At a later stage, the sg. article was created, with the sole function of opposing definiteness to indefiniteness, but bearing the notion of singularity, inherited from its previous stage as a determiner. Finally, the pl. article was created, which came to oppose to the notion of singularity of the article **(h)a(r)* by adding *-k* to this article.

22 We have examples of similar processes (-DV > *-D > -T, where D = voiced stop, T = voiceless stop) in the verbal personal endings, where we find *daki-da-la* 'that I know', but *daki-t* 'I know'. This happens at the synchronic level in many current varieties. But, in addition, in Roncalese the voicedness is kept in spite of the apocope in forms like *ba-daki-d* (**ba-daki-da*), vs. the more widespread *ba-daki-t* 'I know', and in Ultzama (North.-HN), forms like *zai-re* (< **zai-da*), without apocope but with *d* > *r* / V_V are still common, instead of the more widespread *zai-t* (bipersonal verbal auxiliary, 's/he (abs.) - to me (dat.)'). Also in RS (B, 16th century), we find fossils like *dike-a-da* 'I will give you', where *-da* is the marker of 1st p. sg. erg., instead of *-t* (cf. *FHV*: 235-6, Domene 2005: 1st Vol. II, 86).

23 A well-known and often cited example is the family name *Arespacochaga* (already attested in Biscay in 1468), formed by *aretx* 'oak' + *bakotx* 'one' + *-aga*. The presence of 'one' is at odds with the idea of plurality of the suffix *-aga*, which, on the contrary, interpreted as a locative would enable a meaning close to 'place of the solitary oak'.

24 Without making reference to this particular case, Artiagoitia (1990: 332, 340) had already shown that, in Proto-Basque, stops could not stand in syllable-final position. On this impossibility and on the syllable structure in Basque, cf. Lakarra (1995), who shows that, in Proto-Basque, the syllable had to be CV(R)(S), rendering the existence of ***-ag* impossible. For a good summary of Lakarra's theory about the monosyllabic root in Proto-Basque, cf. Chapter 5.

Phases	1	2	3[25]
	- determination - number	+ determination - sg. number	+ determination + sg./pl. number
Morphemes	*-Ø*	Indef.: *-Ø* Def. sg.: *-a*[26]	Indef.: *-Ø* Def.: Sg. *-a* Pl. *-ak*

Figure 6.2. Grammaticalization of the article

According to what we have seen so far, case morphemes were directly attached to the root (Lafon 1999 [1965]: 175, *HB*: 202) before the grammaticalization of the article.[27] Later, the article came about from the 3rd grade demonstrative in order to morphologically mark the definite / indefinite distinction, and finally sg. number. Still later, and in an independent manner, it would be the pl. article's turn, from the previously mentioned *-aga*.

Martínez-Areta (2009b), quite reasonably and following Artiagoitia (1990) and Lakarra (1995), considers that the morpheme *-aga*, from which the grammatical pl. cases arose, could not be derived from *-ag-* (cf. for instance, erg. pl. **-ag-e-k* > **-aek* > *-ak* / *-ek*), for this would be phonotactically very strange. The segmentation *-a-ga* (instead of *-ag-a*) would be more plausible, with the first element corresponding to the article, and the second to a pl. marker (2009b: 251).

In Table 6.6 below, I reproduce an adaptation of Martínez-Areta's table (2009b: 252), which represents his reconstruction of the morphemes prior to the total grammaticalization of the noun paradigm of grammatical cases.[28] His view would imply a simultaneous grammaticalization of case and number (although he (p.c.) now believes that the grammaticalization of number is later than the rise of the case system).[29] Thus, a bare root like *gizon* had a generic value and,

25 As pointed out, in this phase —which is the present one— the penetration of the article as a feature of number has been a gradual process.

26 Or some other morphemes, now disappeared, like those mentioned: *-o*, *-ori*, etc.

27 According to Lafon (1999 [1970]: 171), the fact that in the paradigm of proper place nouns (cf. Table 6.3), which are definite in themselves, we have forms like *Iruri-n* or *Iruri-k* with no trace of the sg. article, bears witness to the previous phase, in which the opposition definite / indefinite did not exist.

28 The author himself indicates that the table shows the elements which take part in this grammaticalization, but that he does not attempt to postulate that this paradigm has ever existed as such in the diachronic axis. It is, then, a theoretical approach.

29 Martínez-Areta (2009a: 84) also believes that the richness of derivational suffixes with abundantial meaning in Basque, like *-(t)z(a)*, *-(t)zu*, *-tsu*, *-d(u)i*, etc. (cf. *dirutza* 'big amount of money', *mukizu* 'squirt (lit. having mucus)', *koipetsu* 'greasy', *pagadi* 'beech

depending on the context, it could refer to a sg. or pl. element.[30] The appearance of the 3rd grade demonstrative (**gizon-ha(r)*), later grammaticalized as an article —marking both definiteness and number—, would initially have served to differentiate definite and indefinite NPs, but since the presence of that demonstrative also denoted singularity, another pluralizing morpheme became necessary, to cover the semantic slot which had thus emerged. This morpheme was **-ga*, after the annexation of which the resulting form was **gizon-ha-ga*. But this morpheme **-ga* is also the ergative marker (or at least homophonous with it).[31] Thus, it could represent both the plural absolutive and the singular or plural ergative, when the case system was not fully developed.[32] According to Martínez-Areta (2009b: 260), this is the explanation for the lack of an indefinite plural. In any case, he is aware of the speculative nature of his paradigm, and assigns it a character of theoretical construct rather than of reconstruction.

Table 6.6. Reconstructed paradigm of grammatical cases hinted at in Martínez-Areta (2009b)

gizon 'man'	Def.		Indef.
	Sg.	Pl.	--
Abs.	*gizon-har*	*gizon-ha-ga*	*gizon*
Erg.	*gizon-har-ga*	*gizon-ha-ga-ga*	*gizon(-e)-ga*
Gen.	*gizon-har-e(n)*[33]	*gizon-ha-ga-e(n)*	*gizon-e(n)*
Dat.	*gizon-har-i*	*gizon-ha-ga-i*	*gizon-i*

6.3.2. The local cases

In the local cases, the features of definiteness and sg. number are not represented by means of the inclusion of the article. Instead, the case marker is directly attached to the root in vowel-ending roots (*etxe-ra*), with epenthesis of an *-e-* in consonant-ending roots (*jolas-e-ra*). Martínez-Areta (2009a: 80) sets out the

wood'), might also be evidence for a period in which there were no number-marking grammaticalized morphemes.

30 Cf. the examples in (3), which show that we can find, still in a recent period, roots without an article with pl. value.

31 The change **-ga* > *-k* of the erg. will be commented on in §6.4.1.1. Castaños-Garay (1979) and Lakarra (2005) believe that the *-k* of the abs. pl. and that of the erg. are originally the same morpheme.

32 In central and western varieties today, *-ak* is still the morpheme for the erg. sg. and pl., as well as for the abs. pl. However, in many of those varieties the plural paradigm has a different accentuation.

33 The *-n* in parenthesis will be explained in §6.4.1.2.

problem in a clear manner, as Michelena (1987 [1971b]) already had. On one hand, the local cases seem to be more recent than the grammatical ones, due to their more lexical nature and their greater phonological content, but on the other hand they seem to preserve traces of a more archaic state, as they do not contain the article. This is precisely part of the evidence behind the assertion that at a certain point in time case morphemes were attached directly to the root, prior to the grammaticalization of the article (Lafon 1999 [1965]: 175, *HB*: 202).

Two further peculiarities of local cases, already touched upon but still worth recalling, are, firstly, the impossibility of their being attached to an animate root, unless we do so by mediation of the morpheme *-(ren)gan(-)* (cf. Table 6.4), and secondly, the appearance of the morpheme *-(e)ta-* in plural and indefinite forms (Martínez-Areta 2009a: 80).

Much has been written about the morpheme *-eta-*. Authors like Trask (1995b) and Urgell (2006) opt for an equivalence with the derivational suffix *-(k)eta*, present in words like *lapurr-eta* 'theft', used to denote an action, and also as a gerund in western varieties: *pents-etan* 'thinking'. As Humboldt and Vinson had already noticed (apud Castaños-Garay 1979: 66), it also appears in place names like *Arrieta*, from *harri* 'stone'. In the light of these place names, Schuchardt attributed a collective meaning to it, and considered it to be of Latin origin, i.e. as derived from the collective *-ēta* (pl. of *-ētum*) (1947 [1923]: 46). Bascologists such as Lafon (1999 [1970]: 168) have shown agreement with Schuchardt on this etymology, but Manterola (2009b: 97 ff.) has recently defended the patrimonial nature of this morpheme, arguing that morphological elements are hardly ever borrowed. Some other problems which he poses to Schuchardt's view are the following: why the pl. should be borrowed but the sg. should not; why the presence of this morpheme conveys not only plural meaning, but also definite value;[34] and why some of the earliest attestations of the morpheme in medieval place names, like *Sansoheta* (Manterola 2008), have aspiration. In fact, the form with aspiration would better explain the variant *-keta* of the derivational suffix mentioned above, which appears in words like *hizketa* 'speech', from *hitz* 'word' (cf. more examples of this alternation in §6.5.2.1).

To sum up, there are two views. Either the suffix *-eta* was borrowed and grammaticalized by entering the declensional paradigm, or, as Manterola defends, the plural paradigm was formed, just like the sg., by starting from the grammaticalization of the demonstrative of the 3rd grade, which employs the patrimonial suffix **-(h)eta* or **-keta* in the local cases. Let us explain the latter view in more detail.

34 In grammatical cases, the article had to specify both definiteness and number, but why should a borrowed morpheme like this do so?

According to Manterola, the whole declensional paradigm, in both local and grammatical cases and in both pl. and sg., would come from the grammaticalization of the 3rd grade demonstrative (not only the sg. of grammatical cases). But, what is more, that demonstrative would have grammaticalized not in its bare form, but already as a declined element. Thus, for instance, if the abs. sg. of the root *mutil* 'boy' is *mutil-a*, which as stated above comes from **mutil ha(r)*, the erg. sg. form *mutil-ak* should not be reconstructed as **mutil(h)a(r) + -k*, but as *mutil ha(r)k*, where the demonstrative itself already has the case marker attached.[35] That said, Manterola suggests the paradigm illustrated in Table 6.7.

Table 6.7. Adaptation of the reconstructed paradigm according to Manterola (2012)

mendi 'mountain'	Sg.		Pl.	
	Form	Origin	Form	Origin
Abs.	*mendi-a*	*mendi ha*	*mendi-ak*	*mendi hak*
Erg.	*mendi-ak*	*mendi hak*	*mendi-ek*	*mendi hek*
Gen.	*mendi-aren*	*mendi haren*	*mendi-en*	*mendi hen*
Dat.	*mendi-ari*	*mendi hari*	*mendi-ei*	*mendi hei*
Instr.	*mendi-az*	*mendi haz*	*mendi-ez*	*mendi hez*
Ines.	*mendi-an*	*mendi han*[36]	*mendi-etan*	*mendi hetan*
All.	*mendi-ra*	*mendi hara*	*mendi-etara*	*mendi hetara*
Abl.	*mendi-tik*	*mendi hatik*	*mendi-etatik*	*mendi hetatik*

Manterola's theory is advantageous from the point of view of economy, since it does not require the postulation of any reconstructed form: as can be observed, nearly all forms with an ungrammaticalized demonstrative are perfectly attested and employed today. It also introduces evident changes in the relative chronology of the grammaticalization of the paradigm. I shall describe this in more detail below, where case forms will be individually handled. In spite of this, this theory also has some drawbacks. Although it is true that the forms above are all attested, they never appear in the same variety. Leaving some secondary developments aside (cf. §7.2.3), Basque dialects do not distinguish between the abs. pl. and the erg. pl. in 3rd grade demonstratives. Thus, the abs. and erg. pl. *hak* and *hek* are two dialectal variants of a demonstrative which represents both the abs. and the erg.

35 Without going so far, Castaños-Garay (1979) had already proposed that the plural, at least as far as the absolutive is concerned, grammaticalized first in the paradigm of demonstratives, and only later passed to the noun declension.

36 Although the 3rd grade singular demonstrative is *hartan* today, Manterola —just like Lafon (1999 [1970]: 168)— considers it an innovation, and believes that the old form is *han*, still preserved as an adverb meaning 'here'.

How can we explain the variant of *-eta-* which appears as *-ta-* in the indefinite local paradigm (cf. e.g. abl. *mendi-ta-tik*)? Trask (*HB*: 204), following Schuchardt (1947 [1923]: 47), thinks that it is the result of a reanalysis. If we take a look at Table 6.1, all grammatical forms present an *-e-* in the plural, with the exception of the abs. This fact would have triggered a reanalysis of the *-e-* of *-eta-* as a marker of plurality, thus giving rise to the form *-ta-* for indefinite forms,[37] since, in the grammatical paradigm, indefiniteness is marked by the absence of an article (Castaños-Garay 1979: 58-9).[38] This is reinforced by the paradigm of consonant-ending roots (cf. Table 6.2), where the *-e-* of *-eta-* continued to be used in indefinite forms, as an epenthesis between two consonants. The disagreement between Trask and Schuchardt is that the latter believed that the number distinction had passed from the local cases to the grammatical ones, not the other way round. Schuchardt considered *-eta* to have had a pluralizing (not locative) semantics, and that once the *-e-ta* segmentation had taken place, *-e-*, interpreted as a plural marker, spread to the grammatical cases —which had already grammaticalized the sg. article— in order to mark the pl. Cf., for instance, erg. sg. *lagun-ak* 'the friend' vs. erg. pl. *lagun-ek* 'the friends'.

Like Manterola, Lakarra (p.c.) argues for an internal explanation for *-eta-*, and also segments *-e-ta-*, but, unlike Schuchardt and others, does not consider it a reanalysis. He believes that *-ta* is in origin an undetermined locative morpheme which is related to the 3rd p. sg. prs. copula *da* 'is', later grammaticalized as an indefinite pl. marker. The first element *-e-*, in turn, would be the 3rd p. pl. morpheme which is found in the verbal system, and which has the variants *-te* and *-de* (cf. B. *dab-e* < **dau-e* '(they) have', *dauka-te* '(they) have', and *dau-de* < **dago-de* '(they) are').[39] If, following Trask (*HB*), the variant with voiced stop is the oldest one, according to Lakarra that pl. marker would be the result of the grammaticalization of the Proto-Basque verbal root **den* 'to finish', found in forms such as *e-ten* 'to interrupt, to section', *lehen* 'in the past' (< **den-en*), and the verbal marker of the past (cf. *zetorr-en* '(s)he came') (Lakarra 2008a: 475-9). This terminative nature would endow the following morpheme *-ta* —in prin-

37 And it would later have analogically spread to demonstratives and pronouns, which today require it when taking a case marker: cf. instr. *zer-ta-z* 'of what', ines. *hone-ta-n* 'in this', etc. (*HB*: 206).

38 Castaños-Garay (1979: 68) shares this view, but does not believe that the grammatical cases have played any role in that alleged "false cut" *-e-ta-*, since neither the erg. nor the dat. present that *-e-* in the plural in all dialects.

39 Gavel (1929) had already established this parallelism, and Castaños-Garay (1979: 95) also discussed it, although the latter did not share it. However, both refer to the *-e-* of the pl. nominal paradigm of grammatical cases, not to that of *-eta-*, as for both authors this morpheme is a suffix borrowed from Latin.

ciple indefinite— with a plural definite semantics. It is possible that this *-(d)e-* is also present in the pl. forms of grammatical cases (Lakarra 2011a: 51).[40] If this were so, the pluralizing morpheme *-e-* might have grammaticalized simultaneously in the local and grammatical cases, against Castaños-Garay's (1979: 77) view that the forms with *-eta-* would be older.

6.4. The origin of the primary cases

As has already been suggested, Basque cases can be classified into different groups. First, we have the core cases (abs., erg., dat.), which —unlike the rest of the cases— are indexed on the corresponding finite verb, and are those which can function as subject, direct object, or indirect object. These three core cases plus the gen. and instr. make up the group of the grammatical cases, which formally present an *-a-* in the sg. and no *-(e)ta-* in the pl. and indef., and semantically are attached to both animates and inanimates. These are opposed to the local cases (ines., all. and abl.), in which the case marker is directly attached to the root in the sg., and *-(e)ta-* appears in the pl. and indef. Besides, these case markers are never —or only exceptionally— attached to animates. Grammatical and local cases together make up the group of primary cases, opposed to the secondary cases (§6.5), which are built upon the primary ones.

The case system cannot be very old, according to Michelena (1987 [1971b]), Trask (*HB*: 246) and Lakarra (2005: 423; 2006b), among others, and it is evident that it has grown in complexity over history. Most case markers are too transparent to be considered old, but even those with scant phonological content, and apparently reduced, have a problem. Although it cannot be discussed here (cf. Trask 1977, Gómez 1994, Gómez & Sainz 1995, Lakarra 2006a, b), there are reasons to believe that Basque may have had an SVO word order, or even VSO or VOS, at some point (see §10.4). This being so, according to Greenberg's universals we should expect the presence of prepositions and a relative scarcity of postpositions in such a period (Lakarra 2006b: 267, 2008a: 455-6). By contrast, the abundance of postpositions is typologically more plausible in SOV languages.[41] Hence, while some cases may have grammaticalized at a stage prior to the SOV order, the latest cases —perhaps the local ones— most probably grammaticalized at a V-final stage.

40 For the variant *-heta* (Manterola 2008), we might conceive **ha-(d)e-ta*.

41 Lakarra (2006a, 2006b) defends the view that Basque started from a situation in which the V was not final, and the language was predominantly prepositional. Thus, the language would be in a state of typological drift towards an SOV order and proliferation of suffixes.

6.4.1. The grammatical cases

6.4.1.1. The ergative[42]

While the absolutive, the covertly marked case, is used for direct objects, object complements (although this function is usually performed by the essive/translative), subject complements, vocatives, and subjects of intransitive verbs, the ergative is employed as the subject of transitive verbs. It shows dialectal variation in its plural form: while continental varieties and some Navarrese ones have *-ek* for the ergative plural, in the rest we find *-ak* (*Lagunek / lagunak ekarri dute* 'The friends have brought it'). Let us recall that Michelena derived the plural article from *-aga* (§6.3.1). The ergative plural marker would be, for this author and for some others (de Rijk 1998 [1981]: 212-3), simply the agglutination of the pl. article and the ergative marker *-k*, which after the apocope of *-a* (*-aga* > **-ag*) would require an epenthesis of *e* to avoid the clash of two consonants (**-ag-e-k*), the intervocalic velar fricative later dropping out, as is usual in Basque.[43] The result is **-aek*, which later phonetically evolved into *-ek* in the east, and into *-ak* in the center and the west.

Trask (*HB*: 200) opts for a second theory, according to which the erg. pl. form would come, exactly like the sg. article, from the 3rd grade demonstrative, but this time from the pl. one. This would allow us to conceive a simultaneous development of the sg. and pl. paradigms, insofar as we would derive the former from the sg. demonstrative and the latter from the pl. demonstrative.[44] Thus, as can be seen in Figure 6.3, in the pl. we would have, just as in the sg., simply the grammaticalization of the demonstrative, with no trace of the morpheme *-aga*.

---

42 Rather than as a purely ergative language, some have defined Basque as an active language, since ergative marking is also used for the subjects of some active, but semantically intransitive, verbs, such as *Patxi-k eskiatu du* 'Patxi has skied', or *Patxi-k saltatu du* 'Patxi has jumped' (Dixon 1994, Etxepare *GB*: 364).

43 Uhlenbeck (1909/10 [1903]: 105) derived the erg. pl. from the agglutination of the pl. article *-ak* and the erg. *-ek*, whence **-akek* > **-aek* > *-ak* / *-ek*. But this would imply that an intervocalic voiceless stop should drop out, which seems very unlikely in Basque. On the other hand, if Lakarra's assumption (§6.3.2) is right, that *-e-* would not be a mere epenthesis. It could be the plural morpheme, grammaticalized from the Proto-Basque root **den*.

44 Plural demonstratives are homophonous in absolutive and ergative forms (cf. §7.2.1), which makes this explanation valid not only for the erg., but also for the abs., and even for the rest of grammatical cases.

gizon ha(r)-k Erg. > *gizona-k* Erg.Sg.
gizon haik Erg. > *gizon (h)aek* > *gizon-ak / -ek* Erg.Pl.

Figure 6.3. Grammaticalization of the erg. sg. and erg. pl. according to Trask

Irigoyen (1981) also favors this view, but believes that the erg. pl. *-ek* would come from a demonstrative **heg-*. However, this reconstruction cannot explain the ergative pl. in *-ak* that we have in the center and west of the entire Basque-speaking territory.

In fact, both approaches have problems explaining the absolutive pl. *-ak*, which is the abs. pl. form in all dialects. The former analysis has to face the previously mentioned semantic problems of the suffix *-ag(a)*, whereas the latter considers the absolutive sg. form to be a reduction of *haik* > *-aik* > *-ak*, disregarding the fact that *haik* is a dialectal and not very widespread form in comparison with the more widespread *hek* and *haiek*, starting from which it would be easy to get to erg. pl. *-ek*, but not so easy to get to the central-western variant erg. pl. *-ak*, nor the general abs. pl. *-ak*.

Lakarra (2005: 442-4) suggests that if, according to the phonotactics of Basque, we propose **-ag* > *-ak* for the pl. article, the erg. marker should be handled in the same way. It does not seem reasonable, then, to stop the reconstruction of the erg. pl. in **-ag-e-k*, since the *-k* of the erg. cannot be anything but a voiced stop unvoiced in final position. Consequently, in Lakarra's analysis, the erg. marker also comes from the already known morpheme **-ga*, the semantics of which is exclusively locative. That morpheme would be the same as the one we find in the local paradigm of animates (cf. Table 6.4), in medial position, and hence with no phonetic modification.

But, how can we link a locative morpheme to an erg. marker? As I have said, Lakarra (2005: 442-444) has suggested that **-ga* is the same morpheme which is used in animate NPs when a local case marker is used (cf. ines. *lagun-a-ga-n* 'in the friend'), and the reason appears to be their high position in the agentivity scale. Animate NPs tend to be agents and local functions are only rarely assigned to them. That is why, according to Lakarra, they would require that morpheme **-ga* to incorporate a local case marker. Since this **-ga*, then, would occur only with animate nouns, it took on an agentive connotation, and, perhaps after turning into *-k*, came to mark the agent in ambiguous sentences where it was not clear which was the agent and which the object (as in 'John kissed Mary'). Finally, *-k* spread by analogy to inanimates, with the same function of marking the agent in opposition to the object.

A potential problem with this view is that it is at odds with the agentivity scale. In fact, if there is an ergative / non-ergative split in noun morphology, we

should expect to find the ergative marking with inanimates, and the non-ergative marking with animates, not the other way round. It is precisely inanimate NPs which, due to their low position in the agentivity scale, require an erg. marker to mark them as agents of an action (Dixon 1994: 83 ff.). One way of avoiding this criticism could be to interpret that, once *-*ga* (> -*k*) had become an agent morpheme, it affected all nouns at the same time, not animates before inanimates. Thus, when *-*ga* appeared only with animates, its meaning was still local, without any agentive connotation.

Another explanation is the one proposed by Martínez-Areta (p.c. and 2012), who considers that the rise of the ergative can be better explained if we interpret the whole process syntagmatically. Following Trask (1977: 209, 1979), he believes that the origin of the ergative is to be searched for in an active reinterpretation of passive sentences, and reconstructs a passive sentence in which the verb would be diathetically unspecified. The passive nature of the sentence would be syntagmatically specified by marking the agent complement with the local marker *-*ga*.[45] Thus, a sentence like **Jone dakus Peru-ga* would mean 'Jone is being seen by Peru'. Later, the agent complement was topicalized, thus moving to occupy the initial position: **Peru-ga Jone dakus* 'By Peru Jone is seen'. Finally, this passive structure was actively reinterpreted, and the agent complement became an agent subject of that active sentence, thus giving rise to the historical ergative.

As for the relative chronology, according to typological principles, ergativity most often arises either in VSO languages or in SOV ones, in both of which SO are next to each other (Dixon 1994: 50). Since most Bascologists believe that Proto-Basque was VO (§10.4), the rise of SOV appears to be a *terminus post quem* for the rise of ergative marking, or perhaps both features —erg. *-*ga* (> -*k*) and SOV— arose as part of the same process.

6.4.1.2. The genitive

If, as explained above in §6.3.1, the plural paradigm of grammatical cases developed according to the first theory —the one which defends the position that they derive from the grammaticalization of the suffix *-aga*, not of the pl. demon-

45 Typologically, it is common to find the patient subject of a passive marked by a locative morpheme. Cf. the locative preposition used in Spanish, as well as its English translation, as in *ha sido vista **por** mi madre* '(s/he) has been seen **by** my mother'.

strative—, the pl. genitive would have undergone the same development as the other grammatical cases: **-ag(a)-e(n)* > **-ag-e(n)* > *-aen*[46] > *-en*.

It is quite commonly accepted that the old genitive was not *-(r)en*, but *-(r)e* (Lafon 1999 [1965]: 173, Michelena 1985 [1972b]: 297), to which the inessive marker *-n* would have been attached. There are several clear pieces of evidence supporting this view. To begin with, this gen. form is the one that appears in pronouns like *gu-re* 'our', *zu-re* 'your (sg.)', etc. It also appears in cases built upon the genitive (§6.5.2), and in fossilized expressions like *are* 'still more', resulting from the union of the article with the genitive (*HB*: 201). It also seems to be attested in medieval proper nouns such as *Orti-re semea* 'Orti's son' (Iratxe, 1125) (Michelena 1987 [1969]: 96), or even *Orci-re çorita* (1110), which was already *Orçi-ren çorita* in 1284 (Irigoyen *DRFLU II*: 177-178).

Michelena (1987 [1970]: 290, 1985 [1972b]: 297) established a parallel between the genitive morpheme *-(r)en* (4a) and the one employed to form relative clauses, *-(e)n* (4b), as well as the past tense marker of the verbal morphology (4c):

(4) a. *Gizon-a-ren* *liburu-a*
man-ART-GEN book-ART.SG
'The man's book'

b. *Gizon-ak* *du-en* *liburu-a*
man-ERG.SG has-RLV book-ART.SG
'The book that the man has'

c. *Zetorr-en* *gizon-a*
came-RLV / PST man-ART.SG
'The man who came / The man came'

As can be seen in (4), the three morphemes are formally very similar,[47] and the function of *-en* seems to be the same in both (4a) and (4b, c): to link a dependent to its head. The difference is that while in (4a) the dependent is an NP, in (4b, c) the dependent is a relative clause (and the *-(e)n* is attached to its finite

46 This phase is residually attested in Old Biscayan and Alavese, as in RS. *iusturi-aen atzean* 'after the thunders'.

47 Another morpheme which can be related to these is the superlative, which today appears as *-en* (e.g. *gizon(-ik) altu-en-a* 'the tallest man'), and which is undoubtedly the same morpheme as the gen. pl. one. This morpheme also shows fossilized traces of an old variant *-e*, as in *ja(u)be* 'owner' (< **jaun-* 'lord' + *-e*) and *hobe* 'best' (< **hon-* 'good' + *-e*) (Lakarra 2006a: 576, 2006c: 581).

verb).[48] Another shared feature is that both the modifying noun in the genitive case and the relative clause precede the phrasal head *liburua*. However, while in old texts, and until recently in far eastern dialects, the order 'phrasal head + relative clause' is productive, the order 'phrasal head + genitive' is rarer.[49] A final difference is that while the gen. *-(r)en* has the variant *-(r)e*, which seems to be older, the *-(e)n* of relative clauses does not have any variant without *-n*. The parallelism between the two linking functions, then, is perhaps to be sought in the simple *-n*, rather than in the whole *-(e)n*.

A final question to clear up in relation to the morphology of the genitive is the origin of the *-r-* which appears in the gen. (and dat.) indef. of vowel-ending roots (Table 6.1), and in the gen. and dat. of proper nouns (Table 6.3). It is usually explained as an epenthetic element, without getting into diachronic considerations. However, following van Eys, Gavel (apud de Rijk 1998 [1969]: 27-28) had already proposed that it came from the grammaticalization of the 3rd grade demonstrative as the article, after an analogical process. If for the abs. sg. we reconstruct **etxe-har*, for the gen. sg. we will reconstruct **etxe-har-e(n)*. From these, after the regular changes we would have abs. sg. **etxe-(h)a(r) > etxe-a* (with the expected loss of *-r* in word-final position), but gen. sg. **etxe-ar-e(n)* (with the expected preservation of *-r-* between vowels). Since in the abs. sg. —probably the most frequently used morphological case in discourse— the article was *-a* after that change, the gen. sg. was reanalyzed as **etxe-ar-e(n)* → *etxe-a-re(n)*, *-r-* now being part of the gen. case marker. After this reinterpretation, the new gen. *-re(n)* was attached to the indef. of vowel-ending roots (→ *etxe-ren*), and to vowel-ending proper nouns (→ *Iruri-ren*).[50]

48 As example (4c) shows, in Basque, verbs in the past are marked by an *-n*, but since the relative finite verbs are also formed by attaching *-(e)n*, past forms with and without relative morphemes are formally identical.

49 Recall that we are talking about the possessive genitive. With the local genitive, the order 'phrasal head + loc. gen.' is much more common (cf. §10.3.3).

50 As is the case for the gen. pl. (cf. §6.4.1.2), western varieties —especially in old texts— also pose particular problems in relation to the gen. sg. Actually, in Old Biscayan we find, in addition to *-aren*, attestations such as RS. *alargun-aen* 'of the widow', Cap. *geure Iaun-een* 'of our Lord', Cap. *Iaungoiko bizi-en-a* 'of the living Lord', Cap. *aiteen* 'of the father' (*aita* 'father'), Bet. *aiten* 'of the father'. According to Castaños-Garay (1957: 65), in all of these, the *-r-* appears to have dropped out. From a common gen. sg. *-aren*, then, we would have *-aren > -aen > -een > -en*, the attestations corresponding to different phases of that development (and the *-a* of *-a*-ending roots being assimilated to the *-e-* of the gen. suffix). A further feature of Old B is the lack of the epenthetic *-r-* that we see in vowel-ending proper nouns, as in Bet. *Eva-en* 'of Eva' or Cap. *no-en* 'of whom?' There are two conceivable explanations: 1) these are also cases of the *-r-* dropping out between vowels, as with *-aren*; and 2) in these cases, the analogical spread of

6.4.1.3. The dative

This case is used mainly as a marker of indirect object. The dat. sg. form comes from the cliticization of the 3rd grade demonstrative (*-har-i* > *-ari*) whereas the *-r-* of the dat. indef. form on vowel-ending roots (*etxe-ri*, cf. Table 6.1) and the one on vowel-ending proper nouns (*Iruri-ri*, cf. Table 6.3) are to be explained as a result of a reanalysis of the type abs. sg. **etxe-har* > *etxe-a*, hence *etxe-ar-i* → *etxe-a-ri* (stem-article-dat.) (cf. §6.4.1.2 for the gen.).

The plural poses special problems. If we start from the same explanation as that of the other grammatical cases, the development would be as follows: **-ag(a)-i* > **-ag-i* > *-ai*. This development would account for the B. G. Al. variant *-ai* (also present in western regions of Navarre), but to the east we still have the variants *-ei* and S. (and adjacent areas) *-er*, for which the origin of the *-e-* would remain unexplained.[51] If we go back to Lakarra's account of the verbal pluralizer *-de* as derived from **den* 'to finish' (cf. §6.3.2), we could interpret that the dative morpheme *-i* has taken this pluralizer *-e-*, thus giving rise to the dat. pl. form *-ei*, in those dialects which have it.

According to Trask (*HB*: 201), *-ei* would be the old form. From this, another form like **lagun-e-r-i* would have been created, *-r-* being an epenthesis analogous to the one that we have already seen in the gen. and dat. indef. of vowel-ending roots. The variant *-er* found from Donibane Garazi to the east would be an innovation with apocope of *-i*. A point in favor of this view is that in the Souletin dialect, where the accent is systematically paroxitone, the dat. pl. has the accent on the last syllable (*lagün-ér* 'to the friends'), which can be explained if we start from **lagün-éri* and assume a later apocope.[52]

The opposite option, according to which the *-r-* may have dropped out between vowels, is not inconceivable, but would be more problematic, since it seems that the loss of intervocalic *-r-* appears to be a recent phenomenon (*FHV*: 311).

To round off the explanation concerning the dative, Lakarra (2008a, 2011a) has recently proposed a lexical source for the pure dative marker —which is *-i* in

-ren to vowel-ending proper nouns did not take place. All in all, however, the overall picture of these phenomena appears to be a western innovation.

51 Unless we postulate, quite *ad hoc*, an assimilation *-ai* > *-ei* (which would not explain *-er* anyway).

52 In fact, the dat. pl. *lagun-eri* 'to the friends' is not only attested, but has been spreading in the continent and northeastern region of Navarre from two or three centuries ago onwards (cf. Pikabea 1993: 45-48). However, as Padilla-Moyano (2011: 41) points out, this does not appear in old texts, at least not in the continent. Therefore, it seems dubious that there is a continuity between the **-eri* proposed by Trask as a source of S. *-ér* after apocope and this modern *-eri*.

the historical period—, deriving it from the grammaticalization of the Proto-Basque root **nin* 'to give', which can be found in finite verbal forms of old texts (e.g. Laz. *in-da-zu*, Dch. *i-da-zu* 'give me!') and fossilized expressions, as well as in auxiliary verbs of tripersonal agreement in central dialects (e.g. *errai-ten d-i-da-zu* 'you say to me', **nin* 'to give' being, among others, the original verbal root of *didazu*, according to Lakarra 2008a: 482). In fact, the verb 'to give' is a common lexical source of dative markers (cf. Heine & Kuteva 2002: 153-4).

6.4.1.4. The instrumental

The instrumental has the function, among others, of marking the object by means of which an action is performed: *mailu-az kolpatu du* '(s)he has hit with the hammer'. But it also holds a great number of other miscellaneous values as in many other languages (Narrog & Ito 2007, Narrog 2009, 2010), which can be temporal, modal, material, etc. Sometimes, the instrumental functions simply as the case governed by another element, whether by a verb, an NP, an adjective, or a participle. It can be also used in antipassive constructions, in which the subject moves from being marked as erg. to being marked as abs., and the direct object, previously in abs., takes the instr. (cf. de Rijk 2002). The purely instrumental function of the instrumental case is actually on the wane, and is increasingly being taken over by the sociative, especially in western areas. Formally, it is commonly *-z*, but the variant *-s* is found in part of the High Navarrese dialect. The latter variant appears to be an innovation, not easy to explain. Perhaps the motivation is phonetic, for there is a phenomenon, not extremely widespread but common in several regions, whereby *z* > *s* / _ T (*FHV*: 282). Examples of this include the place name *Maeztu / Maestu* and the word *uztaia / ustaia* 'horse collar'. The instrumental appears in exactly this context when it takes the local genitive *-ko* (cf. *altzairu-z-ko* 'of steel'). It could thus turn into *-s-*, and later generalize as dorsal in other phonetic contexts.

Whether this is the case or not, this development would be phonetically unusual anyway, for there are not many sibilant fricatives in final position outside the verbal morphology (*HB*: 201). As loanwords like *gorputz* 'body' (< Lat. *corpus*) show, in the final position of old lexemes the sibilants were neutralized in favor of the affricate (*-tz* and *-ts*, seldom *-z* and *-s*). To explain the instr. *-z*, then, it would be reasonable to think that an old vowel which came after the sibilant apocopated. In this case, we would originally have had instr. **-zV*, perhaps **-za*. This form would not need any asterisk if the old instrumental of demonstratives and personal pronouns, e.g. *horr(-e)-za-z*, *zu-za-z* (cf. §7.2.3, §7.3.2), were to be analyzed as the old instrumental suffix *-za* plus its apocopated variant *-z* pleonastically attached.

If *-*za* were really the old form of the instrumental, its resemblance with the derivational suffix of abundance *-(t)za* (as in *diru-tza* 'plenty of money', *elur-tza* 'snowfall', *Elor-tza* 'place of hawthorns') would be striking. A semantic link between the two, however, is more difficult to figure out. Martínez-Areta (p.c.) speculatively suggests that the instr. may have taken on that semantics from contexts in which it is the case required by verbs and nouns of abundantial meaning, as in *poz-ez bete-rik* 'full of happiness', *jende-z gainezka* 'crammed with people', etc.[53]

The morphology of the instr. resembles that of grammatical cases in most aspects. Thus, the instr. sg. is formed upon the article (*etxe-a-z*), whereas the indef. is formed by attaching the pure case marker directly to the root (*etxe-z*). The plural presents an *-e-* (*etxe-e-z*), which makes it similar to erg. pl. *etxe-e-k* and dat. pl. *etxe-e-i*, but in the instr. this *-e-* appears in all dialects. Hence, unlike for the erg. pl. and the dat. pl., for the instr. pl. a unitary explanation appears to be possible, whether that of Lakarra's spread of the pluralizer *-e-* (cf. §6.4.1.3), or any other.[54]

6.4.2. The local cases

Bascologists usually agree that these cases are more recent than the grammatical ones. They have more phonic substance, more semantic content, and typically present polymorphism across both their chronology and geography (Michelena 1985 [1972b]: 303, Trask *HB*: 202).

On the morphological level, as seen above, local cases make use of *-eta-* in the pl., *-ta-* in the indef., and, unlike in grammatical cases, the article *-a-* is not inserted in the sg. At least on the surface, then, the case marker is directly attached to the root, as with grammatical cases before the grammaticalization of

53 An alternative proposal has been made in Igartua & Santazilia (2013). In this proposal, *-(t)za* is regarded as a locative rather than as an abundantial marker in origin. The typological viability of the semantic extension LOCATIVE → INSTRUMENTAL has been defended by a number of authors in recent works. On a theoretical level, it has been defended by Heine *et alii* (1991: 159). Palancar (2002: 125 ff.), in turn, mentions some particular cases of such a process. Likewise, Narrog (2010) compiles several descriptions of concrete cases in Indo-European and Pama-Nyungan languages, among others.

54 In some varieties, the *-(e)ta* morpheme of indef.. and pl. of local cases has spread to the instr. (instr. pl. *etxe-eta-z* vs. *etxe-ez*), but this is a very recent phenomenon. This encroachment of *-(e)ta* on the instr. has most probably been favored by the fact that in personal pronouns and demonstratives the instr. pl. does have *-(e)ta* (*zu-taz* 'of you', *hori-etaz* 'of those'). In old texts, however, we only find *-zaz* for personal pronouns and *-ezaz* or *-ez* for demonstratives (*zu-zaz*, *hori-ez(az)*).

the article, according to Trask (ibid.). The absence of the article in the sg. paradigm, along with the presence of the morpheme *-eta-*, causes this author to believe that the distinction between sg. and pl. that we find in the local cases must be recent and inherited from the grammatical cases. Hence, Trask defends the view that there was a period in which local case markers entailed no number opposition, and were attached directly to the root. The morpheme *-eta-* would simply be a collective suffix which took on pl. value when the local case markers were attached.

6.4.2.1. The inessive

The inessive, formed by the suffix *-n*, can express both spatial and temporal notions. Thus, *mendi-an* means 'on the mountain', but *arratsalde-an* 'in the afternoon'. This case is one of the most studied by Bascologists, and that is not surprising, due to a peculiarity that will be immediately set forth. First, let us take a look at the following examples put forward by Manterola (2009b: 93):

(5) a. *mendi + a = mendia* *mendi-a + n = mendian*
mountain + ART the mountain + INES
'The mountain' 'On the mountain'

b. *lan + a = lana* *lan-a + n = lanEan (**lanan)*
work + ART the work + INES
'The work' 'At work'

If we look at (5a), we notice that the inessive marker seems to be attached to the root with mediation of the article (cf. Lafon 1999 [1970]: 169). This would be the only local case with such mediation, since, as I have said, the rest of them are attached directly to the root. But let us now consider example (5b). Consonant-ending roots generally take the article in the same manner as vowel-ending roots, but not in the inessive. If we attached the inessive marker to the /root-article/ complex, as we find in *mendi-a-n* (5a), we would expect to find ***lan-an*, but this is ungrammatical in any dialect of any period. Instead, we find *lan-ean* (or phonetic evolutions of this). Where does this *-e-* come from? It is undoubtedly an epenthesis. If we compare Tables 6.1 and 6.2, we can observe that, leaving the inessive aside, in order to form the sg. of local cases, when the root ends in a consonant, an epenthetic *-e-* is inserted. This occurs because all local case markers begin with a consonant. The epenthesis of *-e-* prevents the clash of the root-final and the case-initial consonants:

(6) a. *mendi-ra* *mendi-tik*
b. *jolas-e-ra* *jolas-e-tik*

However, the inessive inserts that epenthesis between the root and (what at first sight seems to be) the article. In view of this unexpected phenomenon, Jacobsen (1977) opts for considering this *-a-* to be not the article, but a part of the inessive marker itself, which would be *-an*. In this view, he would later be followed, with some modifications, by de Rijk (1998 [1981]) and Trask (*HB*). Let us look into this explanation in more detail.

As has already been mentioned, the context for the epenthesis is C# _ #C. But here we find C# _ #V instead (cf. *jolas-e-an* 'in the game'). In order to couple the inessive with the structure of the rest of the local cases, Jacobsen suggests that the original form of the inessive was **-Can*. If we start from this case form, the expected inessive of a consonant-ending root would be **jolas-e-Can*, *-e-* being the epenthetic element which prevents the clash between two consonants, exactly as in the other local cases.

In favor of this hypothesis, Jacobsen sets forth a number of arguments. The first is the following. In Biscayan (also in some linguistically Guipuscoan valleys), Alavese and in the area of Burunda (western Navarre), *-a*-ending roots undergo a dissimilation when a case marker beginning with *-a* is attached to them. This dissimilation can be represented in the following way: *-a + -a > -ea*.[55] Thus, for instance, while in the rest of varieties a word like *proba* 'proof' remains *proba* when taking the article (*-a*),[56] in southern Guipuscoa, Biscay and Alava we have *probea*. The same holds for *hanka* 'leg' vs. *hankea* 'the leg', *gona* 'skirt' vs. *gonea* 'the skirt', etc. This occurs with all of the grammatical cases, because they all contain the article *-a-*, but not with the inessive, even if on the surface it also begins with *-a*: *gona* 'skirt', and *gonan* 'in the skirt' (but never ***gonean*).[57] The reason suggested by Jacobsen is that during the period in which the process *-a + -a > -ea* was operative, the inessive marker still maintained the structure **-Can*, which prevented the two *a*-s from being in contact. The original form, then, would have been **gona-Can*, where *-a-* would not be the article. Thus, as occurs with the other local cases, the inessive would not be built upon the article.

55 This form *-ea* can undergo subsequent dialectal changes and evolve into *-ia*, *-ie*, or *-i*.

56 In fact, as already mentioned some varieties display an accentual opposition between the root form and the root plus article form, and in Salazarese the root plus article form undergoes an epenthesis of *-r-* with *-a*-ending roots, i.e. *-a + -a > -ara* (cf. Sal. *alaba* 'daughter' vs. *alabara* 'the daughter').

57 In Old Biscayan and Alavese, this *-a + -an* juncture sometimes appears as *-aan*, as in Laz. *plazaan* 'in the square'.

This explanation can account for the inessive *-ean* of consonant-ending nouns, since *-e-* would be, as in the rest of the local cases, an epenthetic segment preventing the clash of two consonants: *lan*# + #-*Can* > *lan-e-Can* (cf. (5b)). But how can we get from *-*Can* to the historical inessive *-an*? Apparently, by the dropping out of that C in intervocalic position. This is not surprising in Basque phonetics, provided that the C is voiced. In particular, the consonants that can drop out between vowels are *b*, *d*, *g*, *r*, *n* (Jacobsen 1977: 164-5). Of these, de Rijk (1998 [1981]: 217) recalls that *g* is the one that most easily drops out, from a typological perspective, and also in Basque. Furthermore, *g* is also the best candidate for an etymological explanation for *-*Can*. In fact, if we postulate that *-gan* is the source of ines. *-an*, this would not require any asterisk according to Jacobsen, for it is attested in the animate paradigm of local cases (cf. Table 6.4.). If the initial consonant of the morpheme is not lost in this paradigm, it is because here *-gan* is sometimes preceded by the genitive *-ren*, so that the voiced stop is not between vowels (*lagun-aren-gan* 'in the friend'; cf. Table 6.4). We would have, then, an inessive morpheme *-gan*. This theory has been commonly accepted (Trask *HB*: 202-3) until recently, although, as we shall see, there do exist (in my opinion more reasonable) alternatives.

An apparent advantage of the theory initiated by Jacobsen is that, if we assume it, a plausible relative chronology can be established. By the time *-*g*- had dropped out intervocalically, the process *-a* + *-a* > *-ea* was no longer operative. As a result, when it dropped out, the two *a*-s did not dissimilate, but coalesced (**gona-gan* > *gonaan* > *gonan* vs. ***gonean*). The same would have occurred when attaching the inessive marker to the morpheme *-(e)ta* of the pl. and indef. local paradigm: **mendi-(e)ta-gan* > **mendi(e)ta-an* > *mendi(e)tan* vs. ***mendi(e)tean* (ibid.: 164).

As for the origin of *-gan*, Jacobsen (ibid.: 166-7) believes that it was an independent word with locative meaning. According to him, it might have something to do with the common *gain* or its B. variant *gan(e)* 'high place, top'.[58] Whether that lexical origin is correct or not, *-gan* would have been a locative postposition, later grammaticalized as inessive.

De Rijk (1998 [1981]: 217) recalls the locative postpositions used in Basque today:

58 This comparison, however, would be extremely problematic. From a formal point of view, the *-i-* of the common form would be difficult to explain. From a semantic point of view, it would remain unclear as to how a locative marker of such a generic meaning like *-gan* has become as specific as the lexical item just described. Last but not least, the change *-*ain* > *-an* is dialectal and late.

(7) a. *mendi-(a-ren)* *atze / aurre / gain / azpi...-an*
mountain-(ART-GEN) behind / in front / on / under...-INES
'Behind / in front of / on / under ... (lit. of) the mountain'

b. *lagun-a-ren* *atze / aurre / gain / azpi...-an*
friend-ART-GEN behind / in front / on / under...-INES
'Behind / in front of / on / under ... (lit. of) the friend'

As can be seen in (7), the presence of the genitive (with article) is optional on inanimate nouns, but obligatory on animate nouns (cf. ***lagun atzean*).[59] If we consider *-gan* to be a postposition *per se*, we should expect to find the same behavior, and in fact we do find it. With the inanimate nouns of Figure 6.4.a) —the first root being vowel-ending, the second one consonant-ending and the third *a*-ending—, the postposition is attached directly to the root, just as it is optionally in the examples in (7a). In Figure 6.4.b), with animate examples (*abokatu* 'lawyer', *lagun* 'friend', *turista* 'turist'), *-gan* maintains the initial consonant because this is not in intervocalic position.

a. **mendi-gan* > *mendi-an*
**jolas-e-gan* > *jolas-e-an*
**gona-gan* > *gonaan* > *gonan*

b. *abokatu-aren-gan*
lagun-aren-gan
turista-ren-gan[60]

Figure 6.4. Evolution of -V-gan *as opposed to* -(ar)en-gan *according to de Rijk*

According to de Rijk, the forms without the genitive but with an article in the animate paradigm (*abokatuagan*, *lagunagan*, *turistagan*), perfectly possible today, must be later innovations, because they maintain the intervocalic voiced velar fricative. As will be shown below, if we consider the text tradition, this view proves untenable, since the type ines. sg. *lagun-aren-gan* does not appear in the oldest texts.

59 However, Michelena (1985 [1976]: 381) notices an exceptional case where the genitive is not overtly marked: Lç. (*Mk* IX, 24) *haur aitá-k* 'the child's father (erg.)'. As he observes, it is not easy to establish whether this is a simple slip of the regular inflected *haurr-aren aitá-k*, or it represents a remnant of a stage at which such a construction was more productive.

60 In this case, the root-final *-a* and the article have coalesced.

In Jacobsen's view (1977: 166-7), the inessive marker *-n* found in proper nouns of place (cf. Table 6.3) would be secondary. After the loss of *-g-* between vowels, speakers interpreted the *-a-* of the inessive as a marker of sg. article. As a result, *-n* was reinterpreted as the inessive marker proper. Proper nouns of place would use this form "without article" due to their semantic determined nature.

De Rijk (1998 [1981]: 217) disagrees on this point, and believes that the inessive *-n* of proper nouns of place is old. He dismisses the interpretation of *-an* as article + *-n* because there is no trace of the expected forms prior to that reinterpretation (like e.g. ***Iruri-an* 'in Iruri'). In fact, the oldest attestation of a proper place name in the inessive dates back to the 11th century, and in that attestation, as well as ever since then, the inessive marker has been *-n* (with epenthetic *-e-* on consonant-ending roots), with *-a-* never showing up. Moreover, de Rijk notes that to consider *-a-* to be the article would run contrary to the whole argumentation of Jacobsen himself to prove that there was an stop (**-Can*) previous to that *-a-* which, precisely, would prevent that process. I believe that de Rijk is right on this particular point.

By contrast, de Rijk believes that while *-gan* was the case form to mark definite nouns or NPs, *-n* was the one to mark indefinite ones. He splits *-gan*, then, into *-ga* + *-n*, whereby the latter would be the inessive marker and the former a morpheme **-ga* (ibid.: 218).

Where would this **-ga-* have come from? In de Rijk's opinion, it would be the locative morpheme *-aga-*, from which the plural article would have grammaticalized.[61] Hence, in de Rijk's view this suffix has little to do with the notion of plurality, and it would be a morpheme with the function of making locative those nouns which are not locative *per se*, i.e. all nouns except proper nouns of place, which are intrinsically locative. A place name like *Iruri* would take the inessive case with no need to mark its locative quality (*Iruri-n*). By contrast, common nouns, non-locative in themselves, would mark their extrinsic locative nature by means of this morpheme (**etxe-(g)a-n*, **lagun-aren-ga-n*, etc.).

A third view, within the hypothesis that the inessive *-an* comes from *-gan*, is the one set forth by Trask. Let us recall (cf. §6.3.2) that, according to him, in a certain period the local cases lacked morphologized number opposition. By contrast, the grammatical cases did have this opposition, and they would have passed this feature on to the local ones. Trask (*HB*: 204), then, sets out the following chronology:

61 Let us recall that he proposed the evolution *-ag-a* > **-ag* > *-ak*, even if Artiagoitia, Lakarra or Martínez-Areta have argued for the unfeasibility of this segmentation (cf. §6.3.1).

1) Before the grammaticalization of the article and the consequent number distinction, the general locative morpheme was *-n*. By then, the grammatical cases did already have a morphologized number opposition.
2) A new locative is built, by means of the postposition *-gan*, to denote the locative notion on nouns for which this notion is not intrinsic (all except proper nouns of place).
3) On animate NPs, number is marked by the article (+ genitive) (cf. sg. *la-gun-a-ren-gan*). In local cases of animates, the element *-gan* moves on to be reinterpreted as sg., and *-g-* drops out (**mendi-gan* > *mendi-an* 'on the mountain').
4) What in origin was a collective suffix *-eta-* + *-n* moves on to be interpreted as a pl. locative.
5) The *-e-* of *-eta-* is reanalyzed as the same *-e-* as that of the pl. of grammatical cases (cf. *lagun-e-i* 'to the friends'). As a result, *-ta-* turns into a marker of indefinite forms.[62]

There are, however, other views which dispense with the assumption that the ines. *-a-n* comes from **-ga-n*. In §6.3.2 it was shown that Manterola (2009b) argued that article formation consisted in the 3rd grade demonstrative, already declined, being attached to the root. Manterola also casts doubt on the view that the initial consonant of the inessive marker was *g-*. In his view, the consonant of the inessive **-Can* would be the aspiration inherent to the demonstrative. In other words, the morpheme *-han* would consist of the 3rd grade demonstrative plus the inessive morpheme *-n* (*ha-n*). As we have seen, the aspiration later dropped out. In order to explain the inessive of consonant-ending roots, Manterola's analysis can be represented as in Figure 6.5.

a. *lan ha* > *lan-ha* > *lan-a*
b. *lan ha-n* > *lan-e-ha-n* > *lan-e-a-n*

Figure 6.5. Development of C-ending roots according to Manterola

Figure 6.5.a) shows the grammaticalization of the absolutive grammatical case with the word *lan* 'work', while Figure 6.5.b) does the same with the inessive grammatical case. As I have said, the grammaticalization of grammatical cases most probably took place earlier than that of local cases (Michelena 1987 [1981]: 48-9, Trask *HB*: 202). Figure 6.5 displays one difference between the two sets. Grammatical cases on consonant-ending roots (cf. Figure 6.5.a)) are attached in their aspirated form, and later the aspiration drops out, thus giving

62 As we can see, in this account Trask ignores the western variants erg. pl. *lagun-a-k* and dat. pl. *lagun-a-i*, which present *-a-* instead of *-e-*.

rise to the current form. The place names referred to by Manterola (cf. §6.3.1) seem to attest this fact. That is why he dispenses with the asterisk to represent the form *lan-ha*. In local cases (cf. Figure 6.5.b)), by contrast, when *-ha-* is attached to the root, an epenthetic *-e-* has to be inserted to prevent a clash between two consonants (in this case *n-h*), just as in the other local cases, and in the locative itself according to the hypothesis starting from *-gan*.

If Manterola is right, a) and b) of Figure 6.5 correspond to two different grammaticalization periods. a) would have taken place in an earlier period, where the clash *-n-h-* did not trigger epenthesis of *-e-*. b), by contrast, would correspond to a later period, in which for some reason *-n-h-* did trigger epenthesis of *-e-*.[63]

Manterola's account of the inessive *-an* is actually a return to the views held previous to Jacobsen (1977), according to which the *-a* of *-an* is nothing but the article (< 3rd grade demonstrative). It was already defended by Schuchardt (1947 [1923]), Lafon (1999 [1970]) and Castaños-Garay (1979: 139). According to these authors, the *-a-* of the inessive (cf. *iturri-a-n* 'in the fountain') would simply be the article, incorporated into the inessive marker *-n* at a late stage by analogy with the grammatical cases, which do have *-a-*.[64] The main advantage of this approach is that it allows us to dispense with the obligation of reconstructing any morpheme or phoneme. In fact, this account has some other advantages, which I shall later detail. In Manterola's account, however, the grammaticalization of the ines. sg. would have taken place when the element attached to *lan-* still presented the form *ha-*.

The assumption that the *-a-* of ines. sg. *-a-n* is simply the article and not a remnant of a morpheme *-ga-* allows us to dispose of a problem which Jacobsen and his followers, and particularly de Rijk, systematically overlook. It is not easy to understand why the intervocalic *-g-* should systematically drop out in **etxe-gan* > *etxe-an* 'in the house', but instead should always remain untouched

63 Why should this be the case? Manterola (2012) offers two different possible explanations: "An answer for the differences between absolutive and locative definite phrases has already been given by Michelena (1987 [1981a]: 48-49) [...]: he suggests that absolutives attached to the noun earlier than locatives (and plurals), and this difference in time would explain some differences in their respective behaviors". And two paragraphs below: "[...] another possible answer coming to my mind is that both demonstratives, the absolutive *ha* 'that' and the locative *han* 'in that' simply followed different grammaticalization paths, so maybe we don't really need to explain both on the same grounds. This question, anyway, remains open".

64 Such encroachments of morphological features between grammatical and local cases are not unparalleled within the Basque declensional system. As we saw in §6.4.1.4, we have recent attestations of the instr. pl. as *etxe-etaz*, instead of the common *etxe-ez*. Here, the pluralizing *-eta-* of local cases has spread to a grammatical case.

in the sg. forms of local cases with animates like *gizon-a-gan* 'in the man'. To explain this, as we have seen, de Rijk (1999 [1981]: 217) resorts to the argument that this occurred because the variant *gizon-aren-gan* (with *-ga-* attached to the genitive) regularly preserved the *-g-*, and analogically influenced *gizon-a-gan*. However, the *gizon-aren-gan* type appears to be a late innovation. It already appears in Ber. (*Jaungoiko-aren-gan-a* 'towards the Lord'), and, in the plural and the indef., in Pouv. (*gizon-en-gan-a* 'to(wards) the men', *zein emazte-ren-gan-ik?* 'from what woman?'), who also attests the possibility of attaching *-ga(-)* to the abs. pl. form (*guziak gan-a* 'to(wards) all of them').[65] But in both western and eastern old texts of the 16th century we only find the *gizon-a-gan* type.

An apparent advantage of deriving ines. sg. *-an* from **-ga-n* was that it provided an explanation for the fact that in Old Biscayan and Alavese, *-a*-ending roots are not dissimilated when they receive the ines. sg. case (Laz. *ene plaza-an* 'in my square'), unlike in the other grammatical cases (abs. sg. *plazea*, erg. sg. *plazeak*, etc.). The dropping out of *-g-* would have occurred when the dissimilating rule was no longer operative, hence **plaza-gan* > *plaza-an*. However, this hardly entails any advantage over the interpretation of the *-a-* of *-an* as an analogical extension of the article, since within this explanation we can likewise assume that this extension took place when the dissimilating rule no longer applied.

65 In this author, the singular of such forms is still exclusively formed by attaching *-ga-* to the abs. sg.: *Iainko-a-gan-ik* 'from God', *Espiritu Saindu-a-gan-ik* 'from the Holy Spirit', *bere seme-a-gan-a* 'to(wards) his son' etc. The *-(r)en-ga-* structure (*-ga-* upon the genitive), then, appears to have entered the noun declension through the plural and indefinite. It is important to bear in mind that this structure was already found in demonstratives (the 3rd grade forms of which are also 3rd person pronouns), even in the singular: Pouv. *har-en-gan-ik* 'from him', *har-en-gan-a* 'to(wards) him'. This already occurs in Lç. (*har-en-gan-a* 'to(wards) him', etc.), who, in the sg. of the noun declension, only makes use of the *-a-ga-* structure (*-ga-* upon the absolutive, as in *senharr-a-gan-ik* 'from the husband', *Iainko-a-gan-a* 'to(wards) God', etc.). In the plural, this author forms the local cases of animates not by means of *-en-ga-*, but by simply using the usual declension of inanimates: Lç. *gizon-etarik* 'from the men', *Profet-etan* 'in the Prophets', etc. This is most probably the old pattern for the formation of the plural of animates, in which case the *-en-ga-* structure must be an innovation. The following quotation, taken from Creissels & Mounole (2011: 177-8), summarizes how this grammaticalization may have occurred: "Given the tendencies generally observed in phenomena sensitive to animacy hierarchy, it is not surprising that, before generalizing to all animate NPs, the use of *gan* or *baita* in contexts triggering the use of spatial cases was already more systematic with pronouns that with nouns. The fact that the use of *gan* or *baita* was already regular with nouns in the definite singular at a stage of the evolution when direct affixation was still widespread with plural or indefinite nouns is not surprising either. (...) [T]he use of *gan* or *baita* became general in the allative and ablative cases spreading to the locative".

Another advantage of this account is that, unlike the interpretation of *-an* < **-gan*, it provides an explanation for the forms with the inessive in *-n* residually attested in Souletin, such as *etxe-n* 'at home', *gañ-e-n* 'on top', *soñ-e-n* 'on the body', instead of the common *etxe-a-n*, *gañ-e-a-n*, *soñ-e-a-n*, and also for B. *itaurre-n* 'leading oxen', *bat-e-n* 'in one', RS. *orbide-n* 'in the journey', RS. *zaarza gatx-en* 'in bad old age', instead of the corresponding ones with *-e-an*. That these are archaisms not reached by the extension of the article, perhaps because of their adverbial nature, was already defended by Lafon (1999 [1970]: 169). It should also be noted that they belong to marginal dialects, which are typically the last to be reached by innovating tendencies. We have, however, at least two contexts in which the archaic inessive *-n* has not been reached by the extension of the article and which appear in all dialects: 1) the imperfective suffix *-t(z)e-n* (as in west. *esa-te-n*, east. *erra(i)-te-n* 'to say (ipfv.)'), formed upon the verbal nominalizer, and 2) the proper nouns of place (as in *Elorrio-n* 'in Elorrio', *Gasteiz-e-n* 'in Gasteiz', etc.). In both cases, the non-extension of the article is expected. In the former, this is because the imperfective form is not found in a noun paradigm within which *-a* can spread.[66] In the latter, it is because proper nouns are definite *per se*, and hence never take the article.

Those who defend, by contrast, that the *-a-* of ines. sg. *-an* is not the article are obliged to explain these data as innovations. For the pair S. *étxe-n / etxí-an*, Trask (*HB*: 204) ventures to propose an influence from Spanish, where two forms like *en casa* and *en la casa* are differentiated, with the former containing a possessive, even affective meaning, implying 'in my/his/her house'. This explanation, however, is clearly not valid for some of the data above.

A final point to mention concerns the ines. sg. of consonant-ending nouns, which points to *-e-an* in all dialects. Even if Manterola's explanation in Figure 6.5.b) above is plausible, another possibility would be that the extension of the article occurred after the insertion of the epenthetic *-e-*. This would have been inserted to prevent not the clash *lan-* + *-han*, but the clash *lan-* + *-n*, hence → **lan-e-n*, and later → *lan-e-a-n*. This account has the advantage that the stage **lan-e-n* is actually attested in S. *gañ-e-n*, B. *gatx-e-n*, etc. An additional advantage is that we would not require any explanation as to why *-e-* is inserted in ines. *lan-han*, but not in abs. *lan-ha*.

Obviously, the difference between the two options implies a difference in chronology. While in the development of Figure 6.5.b), ines. sg. *lan-e-(h)an*

66 As a matter of fact, in some western varieties *-t(z)ean* is found as the imperfective suffix (as already attested in the Chant of *The Burning of Arrasate-Mondragón*; cf. *TAV*: 81-88). To be more exact, then, we should say, instead, that it is not as expected in this context as it would be within a noun paradigm.

would have arisen simultaneously with the creation of the noun paradigm, the one hinted at in this section would have taken place later, due to an analogical extension of *-a* within a noun paradigm that was already formed.

6.4.2.2. The allative

The allative case, typically employed to mark the target of a movement (cf. *mendi-ra* 'to the mountain'), displays some degree of polymorphism. The most common variant is *-ra*, but in some eastern varieties we find a variant with final stop, *-rat*, and also *-ara*, and in still more eastern varieties, we have *-la(t)* and *-ala(t)*. Contemporary forms such as *bidi-la(t)* (*bide* 'path') come from *bidi-ala(t)*. As is usual in morpho-phonological processes in which a noun declensional suffix is attached to a vowel-ending root, the change *-e-* > *-i-* has been brought about by the juncture *bide-* + *-ala(t)* > *bidi-ala(t)*, with the *-a-* finally dropping out.

Some Bascologists like Gavel (1929: II, 28-9) have stated that there was originally a semantic differentiation between the forms with final stop and those without. Thus, according to Michelena (1979c: 577; with examples from Lç., Ax. and Bp.) and Igartua (1986; with examples from Lç.), *etxe-ra* would be 'to the house', but *etxe-rat* 'towards the house'. Some other authors like Lafitte, Gèze and Txillardegi assign the variant with final stop a more terminative than directional sense, or ascribe a briefer duration to the variant without a stop (apud Igartua 1986). While the issue has not been thoroughly researched, I think that nowadays few or no speakers make that distinction by means of the stop. At least since Schuchardt (1947 [1923]: 56), though, it is clear that it is an independent morpheme.

In the varieties which make use of the forms with *-l-*, these appear in complementary distribution with the forms with simple rhotic. As with the inessive (cf. §6.4.2.1), it is proper nouns of place that behave differently: they form the allative with *-ra(t)*, whereas common names use *-la(t)* (cf. *eliza-lat* 'to the church' vs. *Maule-(r)at* 'to Maule').[67]

De Rijk (1998 [1981]: 219) believes that, if the morpheme **-ga* of the inessive is employed to endow common nouns with locative meaning (cf. §6.4.2.1), the other local cases —e.g. the allative— must have it too. According to him, the

67 It is interesting to note that, in Souletin, the variant with rhotic is usually used for the word *etxe* 'house', even though it is not a proper noun of place. Furthermore, the distinction *etxe-rat* / *etxi-alat* entails the same semantic distinction as the one that I mentioned (§6.4.2.1) in relation to *etxe-n* 'at home' / *etxí-an* 'in the house', in the inessive. Thus, *etxe-rat* means 'homeward', whereas *etxi-alat* means 'to the house'.

-a- that appears as the first segment of some variants, but never on place names, corresponds to this morpheme. To support this view, he puts forward data from Roncalese. In (8a), the locative morpheme is not necessary because we have a place name, and hence the allative case marker is directly attached to the root, with epenthesis of *-e-* to prevent a clash between two consonants. In (8b), the example is a common noun and thus the alleged locative marker would be necessary. The velar fricative drops out in intervocalic position, as usual, and the epenthesis would dissimilate by the closing of *-e-* into *-i-*, before *-a-*. (8c) shows a similar case, this time in Souletin, where the variant of the suffix has *-l-*:

(8) a. *Jerusalem-e-ra*
Jerusalem-EP-ALL
'To Jerusalem'

b. *lurr-i-a-ra* (< **lurr-e-ga-ra*)
earth-EP-LCV-ALL
'To the earth'

c. *oihan-i-a-la*
forest-EP-LCV-ALL
'To the forest'

(8b) and (8c) differ in that (8b) has *-r-*, whereas (8c) has *-l-*. De Rijk (1998 [1981]) explains this difference on phonetic grounds. In Basque, there is a common development whereby *l* > *r* / V _ V (*FHV*: 311-2).[68] It can be noticed in loan words such as *zeru* 'sky' (< Lat. *caelum*), or *borondate* (< Lat. *voluntātem*). Thus, the variant with the rhotic in (8b) might be the result of *-ala* > *-ara*. To explain away the variant in (8c), where the change has not occurred, in that morpheme de Rijk proposes the presence of an old fortis **-L-*, which would naturally evolve into *-l-*. Hence, we would have something like **-(g)aLa*. The origin of that fortis **-L-* might be searched for in the juncture of two consonants, which according to de Rijk are *n* and *l*. Thus, if for the *lurriara* of (8b) he reconstructs **lurr-e-ga-la* > **lurr-e-ga-ra*, for *lurriala* he proposes **lurr-e-ga-n-la* > **lurr-e-ga-La*. That is, the locative marker **-ga-* would have been reinforced by the inessive marker *-n*, and this occurred only dialectally in the varieties with *-l-* in the allative.

68 Michelena even mentions the possibility, with reservations, that all simple rhotics are secondary (ibid.: 327-8).

Trask (*HB*: 205-6) has another view of the facts. Apart from the variants mentioned above, the allative case has another one: *-a*. This appears on place names —which are typically conservative— ending in a consonant, such as *Zarautz-a* 'to Zarautz' and *Irun-a* 'to Irun' (cf. Table 6.3), on adverbials derived from demonstratives such as *hon-a* 'hither', *har-a* 'thither', and after the morpheme *-gan-* of local cases of animates (cf. Table 6.4): *lagun(-aren)-gan-a*, etc. In Trask's opinion, these are conservative contexts, and hence *-ra* is an innovation. When the article was grammaticalized, a form such as *mendi-a* became ambiguous, with *-a* now having two possible interpretations: as the article or as the allative morpheme. In other words, *mendi-a* can be understood as 'the mountain' or as 'to the mountain'. This ambiguity could be removed by inserting an epenthetic *-r-*,[69] but demonstratives and proper nouns of place, which did not take the article, preserved the old form. A factor which may have contributed to the change all. *-a* → *-ra* is the adverbial *hara* 'thither', in origin *har-a*, but perhaps subject to reinterpretation as *ha-ra*. In the forms like *mendi-ala* 'to the mountain' (vs. *mendi-ra*), Trask, unlike de Rijk, does not see the morpheme **-ga-*, and considers them, as I do, to be innovations in which the article has been introduced. As for *-la*, he suggests that it might be an unknown morpheme, secondarily used as allative.

As I have said, in eastern varieties there are variants of the allative with final *-t*. This appears to imply a semantic difference. We have already seen (§6.3.1) that an old stop in word-final position is not possible, and that the ones which exist historically in grammatical morphemes can be derived from a structure DV, after apocope and devoicing of the stop. If this is the case, *-t* can be reconstructed as **-da*.

The origin of this morpheme is a mystery (*HB*: 206). Preposed or postponed, it generates a prolific set of etymological combinations, beginning with the *da-* of the verbal morphology. De Rijk (1995a) reconstructs a morpheme **-dan*, which he postulates to be an old patrimonial morpheme, and assigns it a meaning of 'now'. It would also appear in adverbials like *jada* 'already now', (*ja* 'already then' + **-da* 'now'), and, redundantly and with the inessive *-n*, in adverbials such as *orain-da-n-ik* 'since now' (ibid.: 364). It also appears by analogy with this —but in other contexts due to their loss of meaning— in other adverbials such as *betidanik* 'since always' (ibid.: 369), or *noizdanik* 'since when?' (see §9.4.1.1.1 for further details on this morpheme).

Although the formal equivalence with the *-t* of *-rat* would be perfect, de Rijk does not mention this morpheme in the set of etymological combinations in

69 Just as it occurs in Salazarese with the article when it is attached to *a*-ending roots (*ala-ba-* + *-a* > *alaba-r-a* 'the daughter'; cf. §6.4.2.1).

which **da* takes part, and actually the semantic idea behind an alleged development 'now' → 'towards' is not easy to work out. However, Lakarra (2008a: 459) does include the *-t* of *-rat* within the set of morphemes which could be related to the *da-* of the verbal morphology, although his interpretation of this *da-* is different.

Lakarra's (2006a, 2006b, 2008a) view of the allative is highly interesting. The key for exploring it is his criticism of de Rijk's conception of **da(n)* as an element meaning 'now'. According to him, there are several problems with this assumption. To begin with, on phonetic grounds we cannot have an independent morpheme beginning with **d-* at a very old period, since as Michelena (*FHV*: 257) observed, the natural tendency of **d-* is to become *l-*, so that we would expect ***la-tor*, instead of the historical *da-tor* '(s)he comes, is coming'. On the other hand, on semantic grounds Lakarra (2006a: 587) recalls Lafon's observation that, still in the 16th and 17th centuries, some synthetic verbs beginning with *da-* have a future meaning, which is at odds with de Rijk's interpretation of *da-* as 'nunc'. Hence, following Trask (1977), he ventures to propose that *da-* (~ **-da*) is an old imperfective marker, which only in a recent period has undergone the usual semantic development imperfective → present. As for its lexical source, Lakarra proposes the Proto-Basque root **daR* 'to sit', unproblematically present as the root of the participle *jarri* 'to put, to sit', < **e-daR-i*. As Newman (ed.; 2002) and Heine & Kuteva (2002: 276-278) show, this lexeme is a common source of both locative markers and imperfective (~ progressive) markers.

According to Lakarra, in addition to the preposed **da-* of the verbal morphology, there are two more **da*. All three of them would come from **daR* 'to sit', but the last two would have ended up as postpositions, and finally as suffixes. The first of these *da*, of local meaning, would be the one which grammaticalizes as allative. In order to explain its postpositional nature, and at the same time its phonetic development, Lakarra (2006b: 602) resorts to a morpho-syntactic explanation. Starting from a period in which the verb preceded the object (cf. §10.4.3), we would first have a structure [**da*-V O], which after getting reversed would become [O **da*-V]. At this point the change **d-* > *l-* applied, hence > [O **la*-V], but then the reconstructed morpheme underwent an "excorporation" from the verb, → [O **la* V], and later **la* was incorporated as a postposition, and finally as a suffix to the O, hence → [O-**la* V]. From this construction, the all. sg. form *-ra* would be phonetically regular. The variant S. *-ala* could be explained —as de Rijk did (cf. §6.4.2.2)— by assuming that, in the structure [O-**la* V] which grammaticalized, the O was in the ines. sg., so that the juncture

*-*an-la* would have generated a context from which a fortis arose, > *-alla-* (i.e. *-aLa-*), and hence > *-ala*.[70]

The other **da* which has ended up as a postposition would be that of *etxerat* (< **-ra-da*). This would have grammaticalized as imperfective, just like the *da-* of the verbal morphology, and thus would assign the allative suffix an imperfective, durative semantics. This would explain the difference between *-ra* 'to' and *-rat* 'towards'. How the transformation of a prefix into a suffix has taken place in this case, whether through excorporation and later incorporation, or through "migration" from the word-initial position to the word-final one (as proposed in Lakarra 2006a: 602-606), remains to be clarified. Be it as it may, Lakarra insists that the transformation of a prefix into a suffix is anything but unparalleled in Basque, as happens with *gi-* → *-gi* / *-ki* 'thing, stuff, piece', **de-* / *le-* → *-de* (verbal pluralizer and potential marker, whether different or not in origin), *za-* → *-za* '~ abundantial', and probably some others (cf. Lakarra 2006a: 599-606). The phenomenon is also paralleled in Munda, in several Australian languages, etc.

6.4.2.3. The ablative (and prosecutive)

This local case, which today mainly implies 'from', i.e. 'place from which' in relation to a movement, is also rather polymorphic. Nowadays, the most widespread form is *-tik*, but there also exist other forms, such as *-ti*, *-tika*, *-tikan*, and *-(r)ik*. In its formation, the case marker is directly attached to the root, with no article. From de Rijk's point of view (1998 [1981]: 219), this would be the only local case for which the locative morpheme **-ga* can never be reconstructed for any of the variants, since here the case marker is always attached to the root, with no *-a-* between the two. In Old Biscayan and Alavese texts, another ablative suffix, namely *-rean*, is well attested, together with its pleonastic variant *-reanik*.

Let us begin with the most perceptible difference, which is that between *-rean* and the other morphemes. As Lakarra (1984a) shows, in western texts

70 The power of such a morphosyntactic explanation, however, must be taken with care, since in any case it would not be equally applicable to the whole declensional system. The origin of the cases (whether they come from postpositions, serial verbs, or any other element), then, must be studied individually. Nonetheless, precisely this impossibility of deriving all cases from the same construction can help us to establish a relative chronology of their grammaticalization.

from the 17th century until the disappearance of *-rean* in the 18th century,[71] the two ablative morphemes appear in complementary distribution. Thus, in the oldest texts in western dialects until the 18th century, the ablative case is expressed either by *-rean* or by *-(r)ik*,[72] but never by *-ti(k)*, as the latter does not have an ablative, but a prosecutive meaning. It implies 'through', in relation to a movement. Hence, *mendi-rik* or *mendi-rean* meant 'from the mountain', but *mendi-ti* 'through the mountain'. In fact, in Old Biscayan and Lazarraga's Alavese, the prosecutive —which also inflects for plural under the form *-etati*—[73] is a fully-fledged case different from the ablative.

It is from the 18th century onwards that *-ti(k)* begins to lose its original meaning and encroach on the realm of *-(r)ik*, which will progressively disappear with this sense. We can establish the relative chronology shown in Figure 6.6.

*Stage 1	Stage 2	Stage 3	Stage 4
Abl. *-(r)ik* Pros. *-ti(k)*	Abl. *-rean* / *-(r)ik* Pros. *-ti(k)*	Abl. *-rean* / *-ti(k)* Pros. *-ti(k)*	Abl. *-ti(k)* Pros. *-ti(k)*

Figure 6.6. Chronology of the grammaticalization of the ablative in the west

Stage 1 is not attested, but it is my assumption, partially based on Lakarra (ibid.: 189-190), that *-rean* is an innovation competing with *-(r)ik* from Stage 2 onwards, even if Michelena (1987 [1981]: 44) believed that it was common Basque. Two reasons for these assumptions are the fact that *-(r)ik* appears in place names, which are typically archaizing, and the limited geographical distribution of the morpheme *-rean*.

In 16th and 17th century texts (those belonging to Stage 2), *-rean* is the usual ablative in the west, except on animates, which use *-(gan)ik*. If we leave aside Cap., *-rean* does not appear when the ablative has a temporal meaning either, as in *one-rik aurre-ra* 'from now on', nor when the ablative is attached to a proper noun of place, as in Mic. *Flandes-ik* 'from Flanders'. In the 17th century, a new pleonastic form appears: *-rean-ik*. Already in the 18th century (Stage 3), *-rean* loses ground, and at around the middle of the century it disappears. *-reanik* will continue to survive for one more century.

Also at Stage 3, *-(r)ik* is progressively replaced by the prosecutive morpheme, which keeps its original meaning but, in addition to that, it takes over the

71 From then onwards, it has only survived in archaizing sources such as ballads, which typically maintain structures that have disappeared from the spoken language; *-rean* appears in some of these ballads gathered at the beginning of the 20th century.

72 This is the form still used today for the partitive.

73 As in Cap.: pros. sg. *pensamentu-ti* 'through the thought', pl. *esku-etati* 'through the hands'; Laz.: pros. sg. *bioz-e-ti* 'through the heart', pl. *odei-etati* 'through the clouds'.

ablative function, thus becoming an ablative-prosecutive morpheme. It appears that, on one hand within animate roots, and on the other hand within the plural of any root, *-(r)ik* better withstood the invasion of the prosecutive morpheme. Thus, in 1802, Moguel still preserves forms with *-(r)ik* for these paradigms, while the old prosecutive morpheme, now turned into an ablative-prosecutive one, occupies the whole inanimate sg. paradigm. Gavel (1929: I, 55) believes that the pl. *-etarik* may have arisen as a result of the dissimilation of two stops (*-etatik* > *-etarik*). However, this is hardly plausible, if we consider that in this case the plural appears to have been more conservative, judging from Moguel's attestations. In the other battle, fought between *-rean* and *-(r)ik*, we should further note that *-rean* lost its scope earlier in the plural than in the singular.

Finally, Stage 4 is the one found nowadays in the west. In this phase, the old prosecutive morpheme *-tik*, which over the course of time has become ablative-prosecutive, is the only option available to mark the ablative. As in other regions of the Basque Country, prosecutive notions can also be indicated by means of postpositions after the root in the inessive case, as in *mendi-an zehar* 'through the mountain'.

In the east, the situation is somewhat different. We would originally have had an ablative in *-(r)ik*. When *-ti* enters the declensional system, it does not only acquire a prosecutive value, as in the west, but it also replaces *-(r)ik* in its ablative function, and relegates it —in this case, as in the west— to the plural paradigm (*etxe-etarik* 'from the houses'), place names (S. *Maulé-(r)ik* 'from Maule') and some residual items (Tt. *bekhatu-rik kanpo-an* 'out of sin'). Perhaps the existence of the ablative *-rean* in the west, alongside *-(r)ik*, could slow down the supplantation of *-(r)ik* by *-ti(k)*, which first assumed a prosecutive function (cf. Figure 6.6), contrarily to what happened in the east. It should be recalled that, while in the western RS. *-ti* is only prosecutive and the appearances of *-(r)ik* are ablative (obviously, in addition to those of *-rean*), *-ti* is ablative in old eastern attestations such as Oih. *azi-ti* '~ from the seed' (cf. (9) below) or Tt. *gaiñ-e-ti behe-ra* 'from top to bottom'.

There is great controversy as to the origin and grammaticalization process of these morphemes. Although Uhlenbeck (1909/10 [1903]: 107) postulated an evolution *-tik* > *-ti*, Lafon (1999 [1948]: 199) is right when he states that there are no grounds for such an assumption. Michelena (*FHV*: 236) is of the same opinion as Lafon, and argues that *-(r)ik* has no variant without a final stop (as expected if the alternation between *-tik* and *-ti* were motivated on phonetic grounds).

Schuchardt (1947 [1923]: 51), in turn, proposed that *-tik* was the juncture of *-ti* plus *-(r)ik*. He related the former element to the *-ti* denoting 'place' in words like *mahasti* 'vineyard' (from *mahats* 'grape'), which would come from the suf-

fix *-tegi*, similar in meaning. As for the latter element, he related it to the ergative *-k*, without giving further details. This would not account for the variant *-tika*, already present in Dch. (*zeru-tika* 'from heaven') and Laz. (*Gaztela-tika* 'from Castile'), nor for the variant G. *-tikan*, derived from it.

In fact, the oldest forms appear to be *-ti* (the pure abl.(/pros.) suffix) and *-tika* (*-ti* with the enlargement *-ka*, probably an adverbial suffix, see below), *-tik* being an apocopated variant of the latter. In addition to this morphological argument, *-tik* must be secondary, also on phonological grounds, since a stop in the coda of a syllable is necessarily extrametrical, and hence impossible in Proto-Basque (Artiagoitia 1990).

With regard to the origin of *-ti*, Lafon (1999 [1948]: 200-2) observed that although *-ti* could be phonetically derived from *-tegi*, this derivation was not justified on semantic grounds. He preferred to consider *-ti* to be the same derivational morpheme as that which can be seen on adjectives such as *beldur-ti* 'fearful' (from *beldur* 'fear'), which eventually entered the declensional paradigm. But how? Lafon proposes that this may have occurred after the conversion of an apposition of the subject into a complement of the verb. In fact, he even suggests that this conversion is under way in some attestations, such as in Oih.'s Proverb 44 (9a):

(9) a. *Azi-ti* *bihi-a* (Oih., *Prov.* 44)
seed-'ti' grain-ART
'The grain comes just as the seed is (= just as the seed is, so will the grain be)'

b. *Itsu-a* *da* *bahe-ti* *ez-takusa-n-a* (Oih., *Prov.* 288)
blind-ART is sieve-'ti' not-sees-RLV-ART
'Blind is the one who does not see through the sieve'

A 21st century Basque speaker would translate (9a) simply as 'From the seed (comes) the grain'. But Oihenart's French translation is the following: 'Le grain vient *tel qu'estoit* la semence' [the italics are mine, ES]. What the Souletin author still interprets as an apposition of the subject is apparently taking on predicative prominence, until eventually it has become a verbal complement, semantically ablative (here, of an implicit 'comes'). This is the syntactic path through which the derivational *-ti* would have entered the declensional system, according to Lafon, followed by Michelena (*FHV*: 236), but not by Trask (*HB*: 205). In (9b), by contrast, Oihenart's translation is: 'Celuy la est aveugle, qui ne voit pas au trauers d'vn crible'. In this case, then, *bahe-ti* has already become a verbal complement, now of prosecutive meaning, but Lafon (1999 [1948]: 203) be-

lieves that it must have been, in origin, a derived adjective —just like *beldur-ti*— in apposition to *ez-takusana* (the subject).[74]

It is important to note that, in Lafon's account, once *-ti* had entered the noun paradigm, it adapted to the morpho-phonological processes of this paradigm. Thus, when abl. sg. *-ti* is attached to a consonant-ending root, an *-e-* is inserted: *lurr-e-ti* 'from the earth', just as all. sg. *lurr-e-ra* (as opposed to *beldur-ti*, where *-ti* is derivational).

As for the *-k* present in the variant *-tik*, Lafon believes that it has been inserted into the ablative by analogy with the variant *-(r)ik*. Michelena (*FHV*: 237), by contrast, considers *-k* to be an apocopated variant of *-ka*, this being the same morpheme employed to form iterative adverbs such as *irri-ka* 'laughing' or *korri-ka* 'running'.

If a reasonable source of abl.-pros. *-ti* is the derivational *-ti* (*beldur-ti*, etc.), as the next step we can wonder where this derivational *-ti* might come from. Lakarra (2012b) proposes as a source the verbal root **din* 'to come' (east. *jin* < **e-din* 'to come'), which could also mean 'to become' if we consider the verbal root of subjunctive intransitive auxiliaries *na-din*, *da-din*, etc. Thus, **din* with the meaning 'to come' might have grammaticalized as the partitive/ablative (cf. §6.6.3), whereas **din* with the meaning 'to become' might have grammaticalized as the derivational *-ti*. In fact, the two meanings 'to come' and 'to become' are related in both Romance (cf. Sp. *venir* ~ *devenir*) and Germanic languages (cf. Eng. *to come* ~ *to become*), and also in many more languages beyond European borders.

Lakarra (2012b) also believes that the derivational *-ti* coming from the root **din* would still be attested with a voiced stop, hence in a *fase sparita*, in forms with the structure CVC-*di*, as in *han-di* 'big', *hor-di* 'drunk', etc.,[75] which would literally be 'who has become an animal' and 'who has become a dog' (*hor* was 'dog' in old continental varieties, Lcc. and RS, and most probably also in CB). Furthermore, Michelena (apud Lakarra 2006a: 574 and 2012b) had already noted several decades ago that words like *ber-din* 'equal' (cf. *ber* 'same') and *gor-din* 'raw' (cf. **goR-i* > *gorri* 'red') still have the suffix *-din* with the meaning 'to become'. This can be interpreted as an even more conservative variant of **din* 'to become', or alternatively as a relative clause (*di-* + relative *-n*): *ber-din* 'which has become the same', etc. The (eventually prevailing) variant *-ti* would have been generalized from contexts in which *-d-* devoiced after a sibilant.

74 The original meaning of (9b), then, would have been something like 'The one who does not see "sievous" is blind'.

75 Lakarra here applies a principle taken from Trask (1995b), according to which the variant with voiced stop is earlier than the one with voiceless stop.

The morpheme *-rean* can be analyzed in two ways. It could either be the old form of the genitive *-(r)e* plus the ines. sg. *-an*, as it is according to Michelena (1987 [1981]) and Trask (*HB*: 205; in this case, < *-*ga-n*), or the new form of the genitive *-(r)en*, to which the article has spread in a somewhat unorthodox manner, as Martínez-Areta (p.c.) suggests to me. In either case, the *-r-*, which, attached directly to the root, appears only with proper nouns or in the indef., and in any case only on vowel-ending roots, has morphologized as a part of this ablative (*etxe-rean*, *lurr-e-rean*). It appears, then, that once this abl. *-re(a)(n)* was formed, in order to get into the paradigm it acquired a pl. and an indef. form like the rest of the local cases, by means of the morpheme *-(e)ta*: Cap. *arerio-etarean* 'from the enemies' (he also uses *-etarik*, the distribution being unpredictable; cf. Lakarra 1984a: 178), *zeru-etarean* 'from the heavens' in some verses in the *Doctrina Christiana* by Zubia (1691) (known thanks to the copy made by Lezamis in 1699; cf. *TAV*: 138-139, 162).

In the latter analysis, the *-a-* of *-rean* would be an article of late incorporation, similar to that of the allative in Souletin, and most probably to the ines. *-an* of all dialects. Nonetheless, this ablative is attested as *-ren* in Lcc. (*one-ren* 'from here', *orre-ren* 'from there'; cf. Agud & Michelena 1958: 34), in some posters in verse exhibited in Tolosa in 1619 (*esku-ren* 'from the hand'; cf. *TAV*: 127-131), several times in a text from the beginning of the 17th century from Azkoitia (Guipuscoa), along with *-rean* and *-tik* (cf. Urquizu 2008-09), and even as an archaism in western forms like *liburu-ren bat* 'some book'. The lack of *-a-* in Lcc.'s attestations might be due to their adverbial nature, although there is also Lcc. *a-rean* 'from there', among other cases of *-rean*. The *esku-ren* of Tolosa (in Guipuscoa), by contrast, is within the noun paradigm. If it is not an archaism without an article, it might be a dropping out of *-a-* (*-rean* > *-ren*), or even some kind of metrical license.

All this would imply that the source of the western ablative was the genitive. In fact, identical marking for the genitive and the ablative is not unusual, as the Romance languages show.

6.5. Secondary cases

In this section, I shall discuss some cases which were not included in the paradigms because they are created by being built upon other cases.

6.5.1. Cases built upon the allative

There are two cases which are built upon the allative. One is the directional allative, and the other is the terminative allative. The former can be roughly translated as 'towards', and the latter as 'up to'.

6.5.1.1. The directional allative

The directional allative has already been partially discussed (cf. §6.4.2.2), where it was pointed out that the variant *-rat* of the simple allative may have had precisely this meaning. But there are other options for expressing it. In many areas, this semantic nuance of proximity is expressed by postpositions in the allative case after the bare root, as in *mendi alde-ra* 'towards the mountain' (*alde* 'area'), or in some continental varieties with the postposition *buru-z* in instr. indef. following the root in all. or dat., as in *mendi-ra*/*-ari buru-z* 'towards the mountain' (*buru* 'head'). Another option, actually very recent, is to express the proximate idea by means of the iterative suffix *-ka* (mentioned in §6.4.2.3; cf. *korri-ka* 'running', *irri-ka* 'laughing') attached to the allative case (*mendi-ra-ka* 'towards the mountain').

In the west, the center and in part of Navarre, there is another suffix which covers the same function, and which has the variants *-rantz*, *-rontz*, and *-runtz*. Further variants, such as the one present in *mendi-rantza* 'towards the mountain', seem to be innovations which are formed by pleonastically attaching the allative *-a* to the proximate *-rantz*. Azkue (*MV*: 333), quoting the poet Nicolás Ormaetxea "Orixe" as the source of the idea, suggested that *-rantz* is simply all. *-ra* + *antz* 'face', and recalls that the development 'face' → 'towards' is common in the surrounding Romance languages (cf. Sp. *hacia* < Lat. *faciēs* 'face', or even High Aragonese expressions such as *voy cara al monte* 'I am going towards the mountain').[76] As de Rijk (1998 [1972a]: 121-123) observed, this explanation ignores the variants with *-o-* and *-u-*, and is at odds with the fact that *antz* with the meaning of 'face' is not attested in any dialect.

More recently, Gómez (2005) has reconstructed **-raontz*, from which the three attested variants could be derived.[77] It would be an agglutination of the allative *-ra* plus an old noun **-ontz*, which according to Gómez would be semantically close to Lat. *versus* 'facing → towards', and would have undergone the same grammaticalization. Hence, the Basque reconstruction might be a loan

76 In fact, directional expressions like *cara al monte* are also heard in some areas of Navarre.

77 Gómez mentioned more variants, but the others are clearly secondary.

translation of the Latin *ad versum* (ibid.: 276). The Basque word *ontzi* 'receptacle', in turn, would be a morphological derivation of that **-ontz*.

Lakarra (2008a) suggests that *ontzi* is a loan word (possibly from Lat. *conchea* 'shell-like, concave'), and prefers instead **-ra* + *dontz* (ibid.: 484), or even **-ra* + **-da-non-tz*, as the source of the directional allative. This would be formed by the variant **-rada* (> *-rat*) of the allative (cf. §6.4.2.2) + the Proto-Basque root **non* + the derivational suffix *-tz*. The root **non* would have a meaning of movement which would also be present in *joan* 'to go' (< **e-da-non*) (cf. Lakarra 2006a: 583-4).

6.5.1.2. The terminative allative

The terminative allative appears as *-raño*, *-radino*, *-raino*, *-raindo*, *-d(a)rano*, *-giño*, etc. All of these have final *-o*. De Rijk (1995a: 297-302) believes that the most conservative of these variants is *-raindo*. In fact, forms attested in Navarre such as *bat-e-o* 'at the same time' (from *bat* 'one') could point to the old existence of a consonant which would justify the epenthesis of *-e-*. This consonant would be, according to de Rijk, the *-d-* which is found in *-raindo*, but also in words such as *oraindo* 'until now' (it also appears as *oraino*). **-do* would be, then, an old morpheme to express 'until, up to'. In the rest of the morpheme, in de Rijk's view we would have the word *gain* 'top', which, according to Jacobsen and Trask, is also present in the inessive (cf. §6.4.2.1). Consequently, we would start from the grammaticalization of an independent element **gaindo* (lit. 'up to the top'). As *gain* lost its lexical content, its initial consonant dropped out, thus giving rise to one of the attested secondary cases from the allative: **-ragaindo* > **-raaindo* > *-raindo*. If we start from **-ragaindo*, all the variants mentioned above are easy to explain, except for two.

To explain *-giño*, residually attested in B, we would have to propose that **-gaindo* has been attached directly to the root, with no allative suffix. The second one, *-d(a)rano*, consists of *-da-* + **-ragaindo* > *-raindo* > *-raino* > *-raño* > *-rano*. This *-da-* would be, again, the same archaic morpheme discussed above (cf. §6.4.2.2), with the meaning of 'now', although, by the period of the grammaticalization of the whole suffix, it would already have lost this meaning (ibid.: 362). The desemantization of the morpheme *-da-*, as well as of *-gain-*, further gave rise to a high number of pleonastic forms in many authors, perhaps the most intricate one being Tt. (S, 17th century) *egün-ialadrano* 'until the day'. In this form, we have two already known allatives, *-ala-* and *-ra-*, with which the terminative postposition has agglutinated.

Let us conclude by considering a form which proves problematic for de Rijk's analysis. Dechepare (LN, 1545) has *oraidano*, which like *oraino* means

'until now'. If the nasal comes from the erosion of *gain*, then this form should be reconstructed as **orai-da-gain-do*, where the second morpheme would be *-da-* 'now'. We would have, then, a redundant word consisting of the old and modern forms for 'now', just as we had in the above mentioned adverbial *orain-da-n-ik*. Notwithstanding, in the latter form, de Rijk considered the nasal to be the inessive marker suffixed to *-da-*, not to the erosion of *gain*. In order to explain Dechepare's form, I believe that it would be more economical for him to apply the same procedure, thus starting from something close to **orai-da-n-do*.[78]

Lakarra (2008a: 484 and 2012b) prefers to reconstruct *-ra* + **din* + **non*, with a serial verb construction after the all. suffix. The first would be the monosyllabic root **din* 'to (be)come' (also present in east. *jin* 'to come', < **e-din*), and the second the root **non*, implying both movement and a locative idea. The juncture **-din-non-* would create a fortis environment for the nasal which would prevent it from disappearing. If this were the case, we would be obliged to propose an antihiatic function for the velar of the variant *-ragiño*.

6.5.2. Cases built upon the genitive

This is the case upon which the greatest number of postpositions have grammaticalized. However, we have to divide these subcases into those built upon the old genitive *-(ar)e* (cf. §6.4.1.2), and those built upon the more recent genitive *-(ar)en*, with nasal. The former ones should logically be older, although it is equally possible that those built upon the old genitive have replaced this form with the modern genitive.

6.5.2.1. The sociative

In the first group, built upon the old genitive, we find the sociative *-(ar)e-kin*, as in *gizon-are-kin* 'with the man'. Two variants of the same suffix are *-(ar)e-ki* and *-(ar)e-ki-la(n)*.[79] Schuchardt (1947 [1923]: 61) reconstructed this case suffix as **-(ar)e-kide-n*, which would be the grammaticalization of the postposition *kide* 'companion' with the inessive *-n*, upon the genitive (lit. 'in the company

78 Schuchardt (1947 [1923]: 83) believed that what underlied *oraidano* was: 1) *orai* 'now'; plus 2) the relative form of the verb 'to be' in 3rd p. sg. (*da* 'is' + rlv. *-n*); plus 3) a morpheme meaning 'until, up to', which he reconstructed as **-no* (which in turn could also develop into *-o* if the nasal was intervocalic, cf. *FHV*: 299). Hence, we would have **orai-da-n-no*.

79 The genitive is optional on common nouns, but obligatory on pronouns.

of').[80] This etymology was later accepted by all Bascologists (Lafon 1999 [1965]: 196, Michelena 1985 [1972b]: 297, Trask *HB*: 201).[81] The development would be **-(ar)e-kide-n* > *-(ar)e-kie-n* > *-(ar)e-ki-n*. The intermediate step does not require an asterisk, as it is attested e.g. in E.Cib. *lagun-are-kien* 'with the companion'.

In Trask's opinion (*HB*: 202), the form without the inessive *-(ar)e-ki* would be an innovation by analogy with forms with the local genitive, such as *lagun-eki-ko* (as in *lagun-eki-ko festa* 'the party with the friends'), which have no nasal either. However, I believe, as do Gómez (2005: 275) and Michelena (apud ibid.), that the forms without the inessive —including *lagun-e-ki-ko*— are older. The variant *-(ar)e-ki-la(n)* seems to be an innovation, which added the morpheme *-la* (*HB*: 201), and, in some cases, also the ines. *-n*. The morpheme *-la* could be the same as the one found on pronouns with adverbial meaning like *nola* 'how', *horrela* 'that way', etc.

In western varieties, there is another way to form the sociative. For this, the postposition is not *-ki(n)*, but *-gaz*, and is attached not to the genitive, but to the root in the abs. case. Hence, we have e.g. *gizon-a-gaz* 'with the man' in the singular, but *gizon-akaz* 'with the men' in the plural. The pl. form must be more recent according to Trask (*HB*: 202), and possibly formed by analogy with that of the sg., since in some areas in which the sg. sociative is *-gaz*, the plural is still *-kin*. Morphologically, this pl. sociative is formed by attaching the postposition *-gaz* to the root in abs. pl.: **lagun-ak-gaz* > *lagun-akaz*. As a matter of fact, in Guipuscoa and adjacent regions, where the other sociative —the one with *-kin*— is used, the postposition has also been attached to the abs. pl. of the root: *lagun-a-kin* 'with the friends' (vs. *lagun-e-kin* in eastern varieties). We find the etymological spelling in Guipuscoan authors such as Ub. *Apostolu-ak-kin* 'with the Apostles'.

The western *-gaz* may have been an independent postposition, later grammaticalized. In this case form, Trask (*HB*: 203) identifies, on one hand, the already known morpheme *-gan*, which in Biscayan would have entered the realm of the grammatical cases, and, on the other hand, the instrumental *-z*. Hence, he proposes the following grammaticalization: **-ganez* > **-gahez* > **-gaez* > **-gaiz* > *-gaz*. As we have seen above, Trask believes that *-gan* comes from *gain*. Consequently, he has to resort to this evolution to get rid of the nasal. I agree, however, with de Rijk (cf. §6.4.2.1), that if we assume that *-gan* is nothing but the

80 In Schuchardt's formulation, the genitive *-(ar)e* had lost the *-n*.

81 In fact, Gómez (2005: 275) recalls that this etymology had already been suggested by van Eys (1879: 75-6).

suffix *-ga* + the inessive *-n*, then the reconstruction is as simple as *-ga* plus the instrumental *-z*, as Michelena (1987 [1981]: 44) had already anticipated.

Following Lakarra (cf. §6.4.1.1), I believe that, in origin, the **-ga* of the postposition *-gaz* is the morpheme which ends up becoming ergative, and also the same morpheme which allows animates to take local cases. Morphologically, the instrumental is not a local case (it has *-a-* in the sg., but no *-eta-* in the pl.), but syntactically it is close to them. Bearing this in mind, it appears that the western dialect has formed a sociative suffix, different from the periphrastic one in the rest of the dialects (*-(r)e-kin*), by activating the common instr. *-z* for the local inflection of animates by the same procedure used in the local cases: the insertion of *-ga*. Hence, just as we have ines. *etxe-a-n* 'in the house' but *gizon-a-ga-n* 'in the man', we would have instr. *etxe-a-z* 'with/of the house' but (west.) soc. *gizon-a-ga-z* 'with the man'.

In old western texts and in more recent oral sources, the sociative, both the one in *-gaz* and the one in *-(r)ekin*, sometimes appears with a function which is not sociative, but coordinative, in structures of the type X-*gaz/-(r)ekin* Y 'X and Y'. Here are two examples, one corresponding to Old Biscayan, and the other to Old Guipuscoan: RS. (246) *Ogi-a-gaz hur-a* 'Bread and water (lit. with bread water)', (G, 1705) *asto-arekin idi-a* 'the donkey and the ox (lit. with the donkey the ox)' (in a Christmas song; apud Lakarra 1996a: 193). In addition to Old B and Old G, this coordinative structure also appears in Laz., and there are also traces of it further to the east, as in a Navarrese song gathered by Azkue in his *Cancionero Popular Vasco* (apud Lakarra 2008b: 91), in which we can read *ze-ru-arekin lurr-a / zaure-ak dituzu* 'heaven and earth (lit. with heaven earth) are yours'. Even in the continent, we have found some traces of it, as in the recently published text *K&B* (~ West.-LN, ca. 1750; cf. Padilla-Moyano 2011), and Lakarra (2008b) suggests that it may once have been common to the whole territory.

Similar sociative constructions with coordinative function exist in many languages. Stassen (2000) even makes a binary typological (and geographical) classification into WITH-languages and AND-languages. Lakarra (2008b: 92 ff.) applies Stassen's generalizations to Basque and recalls that languages with coordination of the WITH type usually employ serial verbs, which supports the existence of these in Proto-Basque, and probably also the origin of some case markers from the grammaticalization of such serial verbs. Likewise, the fact that WITH-languages tend to lack case systems and *vice versa* (Stassen 2000: 44) argues in favor of an evolution of Basque in which the coordination of the WITH type is increasingly reduced, and the case system —actually not very old— increasingly larger.

6.5.2.2. The motivative

Among the cases built upon the modern form of the genitive, we have the motivative case. It is commonly represented as *-(aren)gati(k)*, and translated as 'because of', as in *su-a-ren-gatik* 'because of the fire'. Since the postposition has the variant *-gaiti(k)* in western dialects, it was proposed by Michelena (*FHV*: 92), and has been believed since then, that this postposition is a loan translation of Latin expressions such as *hominis causā* 'on behalf of (lit. through the cause of) the man', the second element being in the ablative case. Since *gai* is 'element, thing' in Basque, *-gai-ti(k)* would be the already known ablative-prosecutive morpheme suffixed to *gai*, and the whole postposition added to the lexical item.

In the oldest texts, the postposition is added to the stem plus article in the singular, and directly to the stem on proper nouns. The possibility of adding it to the genitive form, which is optional in the modern standard language, is a later innovation. In the plural, in 16th century texts two main patterns are found: 1) the postposition added to the old form of the genitive (e.g. Dch. *gure gaizki-e-gatik* 'due to our evil deeds', Laz. *an eozen arzai-ae-gati* 'due to the shepherds that were there'), and 2) the postposition added to the abs. pl. form (e.g. RS. *Butroeko meni-ak-gaiti* 'due to the truces of Butroe', Lç. *gure bekatu-ak-gatik* 'on behalf of our sins'). However, the possibility of adding *-gatik* to the modern form of the gen. pl. is already present on some nouns in Lç. (*gizon-en-gatik* 'due to the men'), and in the following century, in Ber. (HN) it is the only option. While the details deserve more research, it seems sound to suggest that the possibility of forming the motivative upon the genitive has entered the noun declension through the plural, and geographically it has spread from the central dialects.

6.5.2.3. The destinative

Another case built upon the modern form of the genitive is the destinative, formally *-(r)en-tzat*, as in *ama-ren-tzat* 'for the mother', and also *-(r)en-dako*, as in *ama-ren-dako* 'idem'. Morphologically, *-(r)en-tzat* is formed by the postposition *-tzat* attached to the modern form of the genitive, and *-(r)en-dako* by the indef. form of the local genitive agglutinated to the modern form of the genitive (the deep form would be *-tako*, which surfaces as *-dako* after a nasal). When *-tzat* or *-tako* are attached directly to the stem, then what we get is the essive/translative case (cf. §6.6.2).

Both forms of this case appear in all dialects from the earliest texts (RS. *idi-en-tzat* 'for the ox', Lç. *bere senharr-aren-tzat* 'for her husband'; Cap. *lagun-en-*

dako 'for the fellow man', Lç. *Iainko-aren-dako* 'for God', etc.), although many varieties have generalized one of them. They are commonly employed with animates, since for inanimates another destinative suffix is used, formed by the juncture of the allative plus the local genitive: *-rako* (cf. §6.6.1), as in RS. *gudu-ra-ko* 'for the war', Lç. *su-ra-ko* 'for the fire'.[82]

Especially in old texts, the destinative function is also sometimes expressed by the simple (either possessive or local) genitive: Tt. *gu-ré oro-ren* 'for all of us', RS. *gibel(-e)-ko* 'for the liver', Oih. *hortz(-e)-ko* 'for the tooth'.

In addition to all these grammaticalized cases or subcases built upon the genitive, there exist some other postpositions which also require the genitive but which have not reached a grammaticalized stage and are consequently written separately, such as *-(ar)en inguruan* 'around', *-(ar)en esperoan* 'waiting for', etc. Some other postpositions require the dative or other cases. Since these have remained at their periphrastic stage, their study would go beyond the scope of noun morphology proper, and consequently they will be omitted here.

6.6. The pseudo-flectional morphemes (or non-cases)

There are a number of morphemes which, although usually considered as part of the declensional paradigm, show special behavior in several respects. In fact, I agree with Trask (*HB*) that in a linguistic analysis they cannot be treated as if they were cases in the normal sense of the word. Consequently, I will analyze them in this separate section.

6.6.1. The local genitive

One of these morphemes is the local genitive *-ko*. The reason for not considering it to be a proper case is that it can be attached to nearly everything, turning the element to which it is attached into an adnominal modifier of the NP head. That element can be, among other things, a case-inflected NP, as in (10a, b). In fact, any case-inflected NP can be turned into an adnominal modifier by *-ko*, unless that case is the ergative or the dative.[83] It can also be any other adverbial, as in

82 In fact, the animate destinative sometimes appears on inanimates as well, as in Mic. *osasun-enzako* 'for the health'.

83 NPs in the inessive case made dependent by *-ko* are not common in modern dialects either. But they do exist in old texts, as in the famous Chant of *The Burning of Arrasate-Mondragón* collected by Garibay in the 16th century (cf. *TAV*: 81 ff.), where we find *oñeztar barru-an-go-ak* 'the Oñacins of inside', in which the inessive adverbial phrase

(10c, d). Moreover, attached to finite verbs plus *-la* they can form dependent clauses, i.e. clauses dependent on an NP head, as in (10e), or adverbial clauses meaning 'because', as in (10f):

(10) a. *lagun-ekin* → *lagun-eki-ko* *festa*
friend-SOC.PL friend-SOC.PL-L.GEN party
'With the friends' 'The party with the friends'

b. *txokolate-z* → *txokolate-z-ko* *tarta*
chocolate-INSTR.INDEF chocolate-INSTR.INDEF-L.GEN cake
'With chocolate' 'The chocolate cake'

c. *gaur* → *gaur-ko* *festa*
today today-L.GEN party
'Today' 'Today's party'

d. *horre-la* → *horre-la-ko* *auto* *bat*
that-ADV.SUFF that-ADV.SUFF-L.GEN car a
'That way' 'A car like that'

e. *Etorri-ko* *dire-la-ko* *sinismen-a*
come-PROSP AUX-CPL-L.GEN belief-ART
'The belief that they will come'

f. *Pozik* *nago,* *etorri-ko* *dire-lako*
happy am come-PROSP AUX-CS.SUFF
'I am happy, because they will come'

In addition to these and some other functions (*HB*: 101-3), the prototypical function of the local genitive is to form adnominal genitive modifiers of NP heads, which —unlike the genitive *-en*— are associated to local notions and typically express origin. In the examples in (11), the *-ko* phrases are modifiers of *jauna* and *Milia*. (11b) is a proper noun attested in a dirge collected by Garibay in the 16th century (cf. *TAV*: 75 ff.). It is most probably archaic, and shows that,

barru-an 'inside' is turned into a dependent of the head *oñeztar* by *-ko*, which becomes *-ko* after *n*. Inessive *-n* + *-ko* structures also survive residually in archaic forms such as S. *etxe-n-ko* 'of home'.

in the past, the modifier in the local genitive case, as in relative clauses (§6.4.1.2), followed the head modified by it.

(11) a. *etxe-ko* *jaun-a*
house-L.GEN lord-ART
'The lord of the house'

b. *Milia* *Lastur-ko*
p.n. (of a woman) p.n. (of a place)-L.GEN
'Milia from Lastur'

Lafon suggested that the suffix *-ko* was derivational, even adjectival, in origin, and only later entered the declension (1999 [1948]: 207, [1965]: 182). In the declensional paradigm, it came to function as a kind of counterpart of the possessive genitive *-en*, with the distribution of semantic scopes already described. He also noted the interesting fact that both suffixes have served, when attached to non-finite verbs, to express future aspect. Broadly speaking, western dialects chose *-ko* (*etorri-ko naiz* 'I will come'), whereas eastern dialects opted for *-en* (*jin-en niz* 'I will come') to express the same function (but see §8.3.3 for the precise distribution of both morphemes).

By contrast, Michelena (1985 [1972b]: 301) noted that, unlike *-en* —which can appear on both animates and inanimates—, *-ko* cannot appear on animate nouns (*seme-aren* vs. ***seme-ko* 'of the son'). As a matter of fact, there is another *-ko* which appears quite abundantly in the scant corpus of the Aquitanian inscriptions (as in Aquit. *SENI-CCO* 'little child', (Soria) *SESEN-CO* 'little bull'), and survives in medieval proper nouns (such as *Enne-co*, perhaps attached to *ene-* 'mine', > Rom. *Íñigo*), and in historical Basque in a few nouns (as in *muti-ko* 'little boy', *zati-ko* 'little piece'). This is clearly a diminutive suffix, hence derivational, and can appear on both animates and inanimates. Taking this into account, as well as the fact that the local genitive does not appear on animates in historical Basque, Michelena seems to suggest that the reason for this gap may be that *-ko* soon lost its function in the declensional paradigm, and was specialized in its derivational nature, diminutive in meaning. Although both elements could be the same, the path of development proposed by Michelena to link them (from declension to derivation) is unexpected according to the principles of typology. Moreover, the derivational *-ko* appears as a productive element in Aquitanian, but is marginal in historical Basque. Thus, it seems to have been on the wane in the prehistoric centuries.

As for its morphological formation, the local genitive functions like the other local cases. It has no article in the sg. (*mendi-ko* 'of the mountain'), but it pre-

sents *-eta* in the pl. (*mendi-eta-ko* 'of the mountains'), and *-ta* in the indefinite of vowel-ending roots (*mendi-ta-ko* 'of certain mountains'). As I have said, *-ko* can be attached to an NP in any case except in the erg., in the dat., and —old attestations and fossilized expressions left aside— in the inessive. Since the inessive is, then, the only local case which presents this restriction, Michelena (1985 [1972b]: 302) proposed that the local genitive was built precisely upon this case. Thus, the underlying form of, say, *mendi-ko*, would be **mendi-an-ko*.[84]

When *-ko* is attached to the allative to form the destinative (cf. §6.3.2), or when it is attached to temporal adverbs (as in (10c)), or to a finite verb plus *-la* to form an adverbial subordinate clause of cause, in some eastern varieties *-ko* appears as *-ko(t)z*. I believe that this added morpheme is the instrumental, which sometimes surfaces as an affricate in word-final position. Not all uses of *-ko*, however, present the allomorph *-ko(t)z*: cf. *etxe-ko* 'of home' vs. ***etxe-ko(t)z*, but *daude-la-ko(tz)* 'because they are', *-lako(tz)* being a complementizer with *-ko(tz)* as a second element.

6.6.2. The essive/translative

Another non-case is the essive/translative *-tzat*, with the possible variant *-tako* in the eastern half of the country. These are used to form object complements, as in the sentence *tontu-tzat hartu nau* '(s)he has taken me for a fool'. It is not considered a case because it can only appear attached directly to the root. Hence, it has neither sg. nor pl.

The etymology of *-tzat* is unclear. If we apply the phonotactic rules that we can assume for Old Basque, the final stop of *-tzat* must have been **-da*. The resulting **-tza-da* could then be split into two morphemes. The first one is formally similar to the instrumental suffix *-za*, perhaps also related to the derivational one of place names, or even to the verbal nominalizer (Lakarra 2008a: 484, and §6.4.1.4), and the second one to the **-da* of the variant *-rat* of the allative, perhaps derived from **daR* 'to sit' (§6.4.2.2). In both morphemes, however, it is very difficult to find a semantic connection with the essive/translative idea.

The variant *-tako* is clearly the local genitive suffix, in its indefinite form. Thus, it appears to have come about as a specialization of the local genitive. 16th century classical authors only use *-tzat*, although the essive/translative *-tako* is attested from at least 1626 in the wedding formula *Nai nauzu zure senarr-etako?* 'Dost thou want me as thy husband?', to be located in the surroundings of Orreaga-Roncesvalles (northern Navarre; cf. *ContTAV*: 105).

84 In fact, the need for an asterisk would be doubtful in view of old attestations such as *barruan-go-ak* 'the ones of inside', mentioned above.

6.6.3. The partitive

Let us conclude this subsection with the partitive *-(r)ik*. As I have already said (§6.4.2.3), the form of the old ablative is the same as that of the partitive. Lafon (1999 [1948]: 206) believed that both the ablative and the partitive came from the same morpheme abl.-ptv. *-(r)ik*, and that the former has become independent as the derivational *-ti* of adjectives incorporated into it. De Rijk (1998 [1996]: 436) and Trask (*HB*: 205) also state that the partitive derives from the old ablative. In fact, marking ablative and partitive notions by means of the same morpheme is a common phenomenon, as can be observed in Romance (Sp. *dame algo de* $_{\text{abl. / ptv.}}$ *esto* 'give me some of this'), Finnish, Kerongo, etc. (Heine & Kuteva 2002: 32-3).

The etymology of this *-(r)ik* is not clear. At first sight, the pure marker appears to be *-ik*, *-r-* being an epenthesis to separate vowel-ending roots from the case marker, just as in the indef. of grammatical cases (with which *-(r)ik* shares precisely the indef. semantics; cf. §6.4.1.2). In turn, *-k* may be compared to the *-k* or *-ka* of abl. *-ti(k)(a)*, perhaps deriving from the iterative suffix *-ka* (§6.4.2.3). *-(r)ika* is actually attested, although perhaps favored *metri causa*, in Port.'s poem of 1610 (*pobreza-rika irten-ika* 'once come out of poverty';[85] cf. *TAV*: 118-122). If *-ka* were the original morpheme, then, the succeeding *-k* would be an apocopated variant. What is left to us is a simple *-i-*. This has been analyzed by Lakarra (2008a: 482) as derivable from **nin* 'to give'. This would imply a grammaticalization path 'to give' → ABL-PTV, which does not seem very compelling, although it has some parallels mentioned by Lakarra himself. Ariztimuño (2012b) prefers to think of **din* 'to come' as the source of the abl.-ptv. *-(r)ik*. The path 'to come' → ABL-PTV seems to be more semantically well-founded, although, as Ariztimuño himself admits, this development has the problem that it would imply that *-r-* is not an epenthesis but a part of the suffix (*-r-* coming from *-l-*, which would in turn come from a **d-* in initial position). If it is a part of the suffix, then we would expect to find it on consonant-ending roots as well (cf. *etxe-ra* 'to the house', *lurr-e-ra* 'to the earth', with the allative), but ***lurr-e-rik* does not exist. We find *lurr-ik* instead.

The difference between the ablative and the partitive, in their historically attested forms, is that the latter does not inflect for number, and thus some authors do not consider it to be a case. Moreover, it fits into the same contexts as the article when the NP is indefinite in both negative (as in (12b)) and interrogative (as in (12c)) clauses, and dialectally in some affirmative contexts (as in (12d);

85 There are actually two *-(r)ika* there. The abl. one is *pobreza-rika*, whereas *irten-ika* is an adverbial built upon the participle *irten* 'to come out' (see de Rijk's analysis below).

cf. de Rijk 1998 [1972b]). Thus, the partitive has been considered an indefinite determiner rather than a case marker by some analysts.[86]

(12) a. *Liburu-a* *du-t*
book-ART have-1SG
'I have the book'

b. *Ez* *du-t* *liburu-rik*
not have-1SG book-PTV
'I have no book'

c. *Liburu-rik* *ba-du-zu?*
book-PTV FOCUS.PTC-have-2SG
'Do you have any book?'

d. *Ba-du-t* *liburu-rik* *(asko / ezer)*
DUMMY.PTC-have-1SG book-PTV (much / something)
'I have books (many books / some books)'

De Rijk (1998 [1996]: 435) identifies two homophonous morphemes of the form *-(r)ik*. One *-(r)ik* is found in predicative complements such as *gizon-a bizi-rik dago* 'the man is alive', or absolute constructions such as *hori esan-ik*, ... 'that said, ...'. The other *-(r)ik* is the partitive morpheme. According to de Rijk, this would be the old ablative, which evolved from being a case marker to its function as an indefinite determiner. Its case marking function is still preserved in certain varieties and in certain contexts. Apart from the cases given in §6.4.2.3, *-(r)ik* with ablative function still survives in fossilized expressions such as *mendi-rik mendi* 'from mountain to mountain' or *urte-rik urte* 'from year to year' (ibid.: 436),[87] and also in superlative NPs such as *emakume-rik ederr-en-a* 'the most beautiful woman' (ibid.: 437).[88] But the most interesting examples are

86 By contrast, an aspect pointed out by Ariztimuño (2012b) which brings the ptv. closer to the case system and away from determiners is that, while determiners can acquire case morphemes, the partitive cannot. On these issues, cf. Etxeberria (2010).

87 This construction is typical in peripheral areas. In more central areas, it is more typical to find the instr., as in *mendi-z mendi* 'from mountain to mountain', the abl. + all. construction, as in *urte-tik urte-ra* 'from year to year', or even the ines. + ines. one, as in Añ. *urte-an urte-an* 'idem'.

88 They can also appear in the bare root form, *emakume ederr-en-a*, or in pl. inessive, always inanimate, *emakume-etan ederr-en-a*.

the following ones, given by de Rijk, which today have virtually disappeared from the language:

(13) a. *Amets-ik* *(asko)* *sortu* *zai-zki-gu*
dream-PTV (many) create AUX-3PL-1PL
'(Plenty of) dreams have arisen to us'

b. *(Ezer)* *gauza* *on-ik* *ni-gan* *ba-da*
(something) thing good-PTV I-INES COND-there.is
'If there is something good in me'

In all of these examples, the partitive always appears with an NP of unspecified value. In (13a), there is an indefinite quantifier, *asko*, which licenses the presence of the partitive on the quantified element, provided that this is of a generic, non-specific nature. In (13b), we find an indefinite quantifier, *ezer*, which licenses the presence of the partitive. In these examples, the partitive fulfills its usual function of linking a non-specific NP to its quantifier (ibid.: 439).[89] The interesting point in these examples is that the quantifier can —and for many current speakers, must— be omitted. We could consider it to have dropped out. After this loss, the functions of the quantifiers were taken over by the partitive, which thus moved from being a link between the quantifier and the quantified, to acquiring a function of indefinite determiner.

From a diachronic point of view, I agree with Ariztimuño that it is misleading to think of one casual partitive and another which is not, as different grammaticalizations of the same original morpheme. I believe that Ariztimuño is also right when he dates the split-off of the ptv. from the abl. —or rather, following Lafon (1999 [1948]), from a ptv.-abl.— to a period prior to the grammaticalization of number and definiteness into the local paradigm, as this would justify the absence of such parameters in the partitive. The ablative would have acquired them at a later stage, since it belonged to the local paradigm.

In Figure 6.7, I reproduce Ariztimuño's proposal. The whole process illustrated there must have been a common phenomenon. The innovation of the abl.-pros. by introducing *-ti* may have been motivated, or at least favored, by the need to distinguish the ptv. from the abl. sg. This abl.-pros. sg. *-ti* arose most

89 The relationship between ablative and partitive is still clear in examples like the following, where the ablative fulfills a partitive function:
Liburu-etatik *asko* *txarr-ak* *dira*
book-ABL.PL many bad-ART.PL are
'Of the books, many are bad'

probably as a reinterpretation of the derivational *-ti* (§6.4.2.3), and came to replace abl. sg. *-(r)ik*, pushing *t*-less ablatives into archaizing contexts (the plural, place names, etc.).

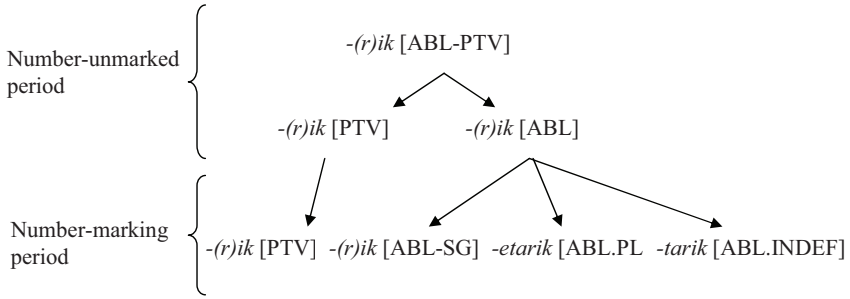


Figure 6.7. Grammaticalization of the ablative / partitive according to Ariztimuño (2012b)

6.7. Conclusion

In this chapter, the rise of definiteness and number as grammaticalized categories has been described, as well as that of the cases within the case paradigm, which, as we have seen, have been discussed for many years. The idea that emerges is that these grammaticalization processes correspond to a transition from a more isolating stage to a more agglutinating one. Even so, the grammaticalization of the declension system as a whole is not homogenous, since morphemes have arisen in different periods, through diverse and unsystematic paths.

Perhaps it is time to reconsider, from the current state of research, the grammaticalization of each morpheme, and jump from a case-by-case study of the grammaticalization of each morpheme to a more paradigmatic vision. After searching for explanations for individual morphemes, and discussing the ones proposed so far, we should further pursue a global interpretation of the rise of the paradigm. I believe that it is viable to establish a relative chronology of the Basque case paradigm as a whole. In this chronology, several phases along the path isolation → agglutination → fusion could be sorted out, once we have synchronically identified the different groups, such as the local cases, the plural of which, formed by means of the morpheme *-eta-*, is neatly agglutinating, unlike the plural of grammatical cases. An element which could help is the study of the chronology of intervocalic epenthetic consonants, which appear only in some cases (cf. gen. indef. *etxe-r-en* vs. pl. *etxe-en*), and hence, in certain periods.

A promising factor in this enterprise is that we own progressively more and more typological information as material for comparison, which allows us to prioritize the hypotheses which best fit into the most frequently attested grammaticalization paths. In spite of recent contributions such as Lakarra (2005, 2006a, 2006b, etc.), Manterola (2009b) and Martínez-Areta (2009a, b), which aim at reconstructions of the whole paradigm rather than of concrete cases, and employ typological comparison in their research, this aspect has not been sufficiently exploited. Comparing the Basque noun declension to that of other languages, and looking for similar patterns and paths of development is a method which must be taken advantage of, as far as linguistic typology will allow us.

7. Demonstratives and Personal Pronouns

Mikel Martínez-Areta
(University of the Basque Country, UPV/EHU)

7.1. Introduction*

In this chapter, I present a section of morphology which, in the handbooks of many languages, is usually sandwiched between the noun and verb morphologies, although it actually belongs to the former, or at least to its (sometimes irregular) borders. As a matter of fact, in this section one should ideally include not only demonstratives and personal pronouns, but also interrogative and indefinite pronouns, and perhaps also numerals and other grammatical items, but this would make the whole picture rather irregular and inconsistent for the general purposes of this work. Hence, I prefer to stick to these two realms, since they are both well-defined synchronically, and are diachronically the most interesting (further details are found in *HB*: 96-99, 196-201, and Trask *GB*: 118-136, 150-160).

7.2. Demonstratives

7.2.1. The paradigms of the standard language

Most varieties of Basque, and most probably also the common language, have three series of demonstratives, which correspond to three deictic grades: 'this', 'that', and 'that yonder'. All of these demonstratives can function as both pronouns and as postponed determiners. Each of the grades has a singular and a

* This chapter has been written within the framework of three research projects: 1) The Research Project "Monumenta Linguae Vasconum III-IV: historia, crítica y edición de textos vascos" [= FFI 2008-04516, FFI2012-37696], led by Joseba Lakarra and funded by the Ministry of Education and Science of the Spanish Government; 2) The Research Project "Historia de la Lengua Vasca y Lingüística Histórico-Comparada" (HLMV-LHC), Ref. GIC 10/83, IT 486-10, led by Joaquín Gorrochategui and funded by the Government of the Basque Autonomous Community; and 3) The Training and Research Unit "Hizkuntzalaritza Teorikoa eta Diakronikoa: Gramatika Unibertsala, Hizkuntza Indoeuroparrak eta Euskara" (UFI 11/14), funded by the University of the Basque Country (UPV/EHU). I thank Joseba Lakarra, Borja Ariztimuño, Blanca Urgell and Manu Padilla-Moyano for their comments on a previous version of this chapter. All errors are mine.

plural, just like the noun declension, but no indefinite, since deictic items are intrinsically definite. For the same reason, in the case axis they have no partitive case. They also lack the essive/translative case, but they do have forms for the rest of the cases.

Since demonstratives have a high frequency in discourse, they tend to undergo intensive phonetic erosion. This has brought about a huge diversity in their forms throughout the different dialects. Perhaps the simplest way of getting an overall picture of Basque demonstratives is to present the paradigms of Standard Basque or *euskara batua*. These are very conservative in form. In fact, these paradigms do not exist as such in any contemporary variety, for in current speech such reductions as *hauek* > *ok* have taken place. The paradigms are given below in Table 7.1.[1]

Table 7.1. Paradigms of 1st, 2nd and 3rd grade deictic demonstratives in Standard Basque

		***hau* 'this'**	***hori* 'that'**	***hura* 'that yonder'**
Sg.	Abs.	*hau*	*hori*	*hura*
	Erg.	*honek*	*horrek*	*hark*
	Gen.	*honen*	*horren*	*haren*
	Dat.	*honi*	*horri*	*hari*
	Loc.-Gen.	*honetako*	*horretako*	*hartako*
	Ines.	*honetan*	*horretan*	*hartan*
	All.	*honetara*	*horretara*	*hartara*
	Abl.	*honetatik*	*horretatik*	*hartatik*
	Instr.	*honetaz*	*horretaz*	*hartaz*
Pl.	Abs.	*hauek*	*horiek*	*haiek*
	Erg.	*hauek*	*horiek*	*haiek*
	Gen.	*hauen*	*horien*	*haien*
	Dat.	*hauei*	*horiei*	*haiei*
	Loc.-Gen.	*hauetako*	*horietako*	*haietako*
	Ines.	*hauetan*	*horietan*	*haietan*
	All.	*hauetara*	*horietara*	*haietara*
	Abl.	*hauetatik*	*horietatik*	*haietatik*
	Instr.	*hauetaz*	*horietaz*	*haietaz*

1 These paradigms are not represented in their extensive version, since details about the secondary cases do not belong in this chapter. Thus, the cases which derive from the allative, such as the terminative (*honetara-ino* 'up to this') and the directional (*honetara-ntz* 'towards this'), as well as the dialectal variants of this case (east. *hon-etarat*, etc.), are not included.

7.2.2. Differences across dialects and historical developments

The diversity of forms, both dialectal and diachronic, is more marked in the plural than in the singular forms. An important feature of some eastern varieties, however, affects all numbers and cases alike. In Aescoan and a part of Southern High Navarrese (see *OEH*),[2] the characteristic *h-* of demonstratives appears as *g-* (*gau* 'this', *gori* 'that', etc.), and, in Salazarese and Roncalese, as *k-* (*kau(r)*, *kori*, etc.). These stops are sometimes considered archaic (see *HB*: 181), but actually the phenomenon is not a regular one,[3] and as Trask himself recalls, the article *-a* (presumably < **har*; cf. §7.2.4) is stopless in these varieties (cf. also §2.2.2.9, in footnote, on the possibility of them not being archaic).

As for the *h-*, present in all forms of Table 7.1, it is merely graphic in the standard language, and even as a grapheme it does not appear in the literary tradition of the peninsular dialects. In fact, its pronunciation is limited to the varieties in which aspiration has been maintained (§4.4.3.1.1).

In the singular, the 1st grade *hau* appears as *haur* in Low Navarrese (which also has *hau*) and Souletin, and as *kaur* in Roncalese. The final rhotic was more widespread in 16th-17th century texts, but has been on the wane since then (see *OEH*: 'hau'). The 3rd grade *hura* does not appear in Biscayan, which has *a* instead. In the cases other than abs., the three deictics have a different root in all dialects: 1st grade *hon-*, 2nd grade *horr-*, 3rd grade *har-*. The root *hon-* closes to *hun-* in the continent, although *hon-* also appears in Dch. and some Old S. authors. The gen. and dat. forms have no special characteristics, although the simple rhotic of *har-en* sometimes drops in the west, as e.g. in Laz. Cap. *aen* (both have *ar-en*, too), while it becomes multiple in HN. *arr-en*.

In the local cases, dialectal and historical diversity basically concerns the case forms, which will not be discussed here. An important development, however, is the tendency of 1st and 2nd grade demonstratives with three or more syllables to drop the second syllable, that is to syncopate, in some dialects. At least

2 *OEH* = *Orotariko Euskal Hiztegia*, or *General Dictionary of Basque*. It corresponds to Euskaltzaindia (1987-2005) in the References and, as indicated there, it is also available on the Internet.

3 In the Roncalese texts collected by Bonaparte, as a 3rd grade demonstrative *kura* always appears as *k-*, but as a 3rd p. pronoun, it appears twice as *kura*, and four times as *(h)ura* (see Pagola *et alii* (eds.) 1992-1999: Roncalese: 263, 267). In L.Elc. (South.-HN), adverbs and pronouns have apparently no initial stop (*or* 'there', *ori* 'that, you (formal)'), but determiners do (*mai-goi* 'that table'). Even admitting this for L.Elc., with regard to the problem of these forms in general in Aesc, Sal and R, Lafon (1999 [1955]: 720-721) believed that a given demonstrative was stopless only when it was anaphoric (not deictic). It is not easy to determine which one of these was the original distribution between forms with and without stop, if there was one (cf. Camino 1997: 410-414).

as far as the former is concerned (see *OEH*: 'hau'), in the continental Basque Country, forms of the *huntan* type alternate with those of the *hunetan* type (both < *hon-etan* 'in this') from the mid-17th century, and they are predominant by the beginning of the 19th century. In the peninsular Basque Country, the breakthrough of the *ontan* type took place at a later date. In Guipuscoan, the *ontan* type appears and its frequency increases over the course of the 19th century, and by the 20th century, it is dominant over the *onetan* type. In Biscayan, by contrast, *on-etan* has been the main form in the historical period.[4] As for the 2nd grade demonstrative, the *CA*[5] gives only Volt. *soku hor-tan* 'on that corner' and E.Cib. *hor-tarakotz* 'because of that' in the first half of the 17th century, but in the second half, syncopated forms already appear in plenty of continental authors from Labourdin (Pouv., Gç.) to Souletin (Oih., Tt., Bp.). Among peninsular authors, the first syncopated attestations which I have found are from the second half of the 19th century for Biscayan, and from the second half of the 18th century for Guipuscoan and for Southern High Navarrese.

Plural forms are more diverse in their structure. To begin with, the 1st and 2nd grade B. (but not Al. —at least not Laz., although Bet. has *onek* 'these here'— or G.) forms are based on the root which, in the singular, forms the cases other than the abs. (1st grade *hon-*, 2nd grade *horr-*). In the rest of the dialects, the pl. root is that of the abs. in the sg. (1st grade *hau-*, 2nd grade *hori-*). Table 7.2 is a simplified version of this dialectal alternation.

The forms in column b of Table 7.2 are, apart from those of the standard language (see Table 7.1), a diachronic projection, as most of the forms attested to the east of Biscayan derive phonetically from them. This evolution has produced an extremely diverse array of forms, and, in some cases, syncretism among deictic grades. Thus, in the work of several authors —not from any particular region— the 1st grade pl. forms undergo reductions like abs.-erg. *hauek* >

4 In High Navarrese, *ontan* already appears twice in an award-winning poem in Iruñea in 1609, although in the same poem, *onetan* also appears twice and *onetako* once (cf. *TAV*: 111-118). Even if these cases of *ontan* might have been favored *metri causa*, *ontan* and other syncopated forms are not uncommon in 17th and 18th century HN. texts (see HN. texts in Satrústegui 1987). For many of these texts, it is difficult to locate which variety they actually reflect (see Camino 2004: 33-52). To pick just one concrete example, in a text from 1751 whose language Camino (2004: 48) associates with Goñerri (several kilometers to the west of Pamplona-Iruñea; see Satrústegui 1987: 85-91), I find seven cases of *ontan* and three of *ontako*, with no unsyncopated counterexamples.

5 *CA* = *Corpus Arakatzailea*, or *Scrutinizing Corpus*. This is a computerized database of the main works written in Basque since the 16th century, which can be consulted according to types of text, dialects and periods. It does not contain all works, and it may contain some typographic mistakes, but to get a general picture of a given problem it provides useful and reliable information (see *Corpus Arakatzailea* in the References).

hoiek, *hok*, gen. *hauen* > *hoien*, and so on.[6] Since the resulting forms are phonetically quite close to the 2nd grade pl. demonstratives (abs.-erg. *hori-ek*, gen. *hori-en*...), either due to loss of intervocalic *-r-* or to simple morphological attraction, in some authors or varieties of different periods the two series appear to have fused together. Thus, already in Dch. we find abs.-erg. *hok*, *hoiek*, gen. *hoien*, *huien*, and some other *hoi-* forms, but no corresponding *hori-* form. That is, there are 1st grade pl. demonstratives but no 2nd grade ones (cf. *EMH*: 14). In other cases, in spite of this phonetic evolution, both series have remained separate, as in Lar. abs.-erg. *óiek* 'these' but abs.-erg. *oríek* 'those' (Irigoyen 1981: 370). The same is true for the Souletin Bp., which has 1st grade abs.-erg. *hoiek*, but 2nd grade abs. *horik*, erg. *horiek* (apud ibid.: 371). In 20th century Baztan, the 2nd grade form had lost the *-r-* (abs. erg. *óiek*), but since the 1st grade form was *ok* (< *hau(e)k*), there was no syncretism either (apud ibid.: 373).[7]

Table 7.2. 1st and 2nd grade plural demonstratives in Biscayan and other dialects

a.	**Biscayan**	b.	**Other dialects (diachronic projection)**
	ón-e(e)k 'these (abs., erg.)'		*hau-ek*
	ón-e(e)n 'of these'		*hau-en*
	ón-e(r)i 'to these'		*hau-ei*
	ón-e(e)tan 'in these'		*hau-etan*
	órr-e(e)k 'those (abs., erg.)'		*hori-ek*
	órr-e(e)n 'of those'		*hori-en*
	órr-e(r)i 'to those'		*hori-ei*
	órr-e(e)tan 'in those'		*hori-etan*

3rd grade pl. forms pose special problems. Many non-western forms are built upon a root *ha-* which takes an epenthetic *-i-* (phonetically a glide) when case forms are affixed: Dch. abs.-erg. *haiek* (along with abs.-erg. *hek*), gen. *haien*, *heien*, abl. *haietarik*, Lç. *haien* (but dat. *haei*!), Oih. erg. *heiek*,[8] L.Elc. abs.-erg. *(g)aiek*. However, in the west (B and Al, and in some cases G; cf. *OEH*: 'haiek'), the oldest forms are found without *-i-*, as in Cap. abs.-erg. *aek*, gen. *ae(e)n*, ines. *aetan*, *aekgatik* 'because of those yonder', Laz. abs.-erg. *aek*, gen. *aen*, etc. This implies that the epenthetic *-i-* has not been inserted in western forms, even though they start from the same morphological juncture (*ha-* + abs.-

6 Oihenart (1992 [1638/1656]: 60 ff., apud Irigoyen 1981: 366) indicates that monophthongized forms like *hoiek* for *hauek*, etc., are popular ("pro *hauëc*, *hauën*, *hauëy*, *hauës*, solet Vulgus *hoiec*, *hoien*, *hoiey*, *hoies*, pronuntiare").

7 High Navarrese data are too diverse to be taken up here, but it should be mentioned that in Southern High Navarrese we have *(g)ebek* for *hau-ek* in the 1st grade, with vowel assimilation and spirantization of the glide.

8 For eastern data in what remains of the section, see *EMH*: 13-16 and *OEH*.

erg. *-ek*, gen. *-en*, ines. *-etan*, etc.). In later Biscayan authors, another 3rd grade pl. form, *ár-eek*, spreads, which is dominant by the 19th century. It appears to be formed by attaching the abs.-erg. pl. *-eek* to the 3rd grade root, which in the sg. is *ar-*. This *ár-eek* may have reduced to *ár-ek*, as in Lekeitio and Markina, or to *ar-eik* as in Aramaio.[9]

A final set of 3rd grade pl. forms which should not be left unmentioned is the one formed —at least synchronically— upon the root *he-*, found in continental authors. Thus, Dechepare has abs.-erg. *hek*, along with abs.-erg. *haiek* (and one case of abs. *heiek*), and also ines. *hetan*; Leiçarraga has abs.-erg. *hek*, soc. *hekin*, ines. *hetan*; Axular has abs.-erg. *hek*, ines. *hetan*; and Oihenart has abs. *hek*, ines. *hetan*. The diachronic interpretation of these *he-* forms is controversial. Azkarate & Altuna (*EMH*: 18) consider *hek* to be a reduction of *ha(i)ek*, whereas Manterola (2008) prefers to consider *ha(i)ek* a pluralization of the 3rd grade deictic root (*ha- + (h)ek > ha(i)ek)*).

7.2.3. Peculiarities of case forms as compared to the noun declension

To begin with the sg., the abs. form of the three deictic elements is the simple root (*hau(r)*, *hori*, *hura*; for details about the latter, see §7.2.4). All the other cases are formed upon a somewhat different root. The erg., gen. and dat. are constructed by suffixing the pure case form (*-k*, *-en*, *-i*, respectively) to that root, just as in the indef. of the noun declension and on proper nouns.

In the erg., the *-e-* of *hon-e-k* and *horr-e-k* is epenthetic, to prevent the consonant clashes ***hon-k*, ***hor-k*, as is usual in the noun morphology. By contrast, erg. *hark* has no epenthetic vowel in most varieties and in the oldest texts. Perhaps the reason for this is that its rhotic is simple, not multiple.[10]

The gen. sg. of demonstratives is not the archaic *-e* which appears on personal pronouns, but the more modern *-en*, as in the noun declension.

Local cases also have, in the sg. —that is, unlike noun declension—, an element *-eta* which appears as *-tan* in 3rd grade forms (e.g. *ha(r)-tan* 'in that yonder'), and also in some 1st and 2nd grade forms after syncope, depending on the dialect and the period (see §7.2.2). This *-eta* appears as a derivational suffix on

9 From this point onwards, the contemporary Biscayan data for demonstratives are taken from: Hualde & Elordieta, G. & Elordieta, A. (1994: 97) for Lekeitio; Rollo (2006 [1925]: 31) for Markina; Ormaetxea (2002: 70) for Aramaio; Gaminde (1997: 50) for Gatika; and Salazar (2001: 42) for Ugao-Miravalles.

10 In fact, both the possibility of dropping the rhotic (e.g. RS. erg. sg. *ak*) and of inserting an epenthetic *-e-* (e.g. Mog. L.Elc. *ar-e-k*) are diversely attested, but the common form is most probably *har-k* (Laz. *ar-k*, Lç. Dch. *har-k*).

place names, implying abundance (*(H)arri-eta* 'stones', *Art-eta* 'holm oaks'), or location in some cases (*Zubi-zarr-eta* 'the place of the old bridge'). In the noun declension, *-eta* appears in the pl. and the indef., although in the latter the underlying morpheme is *-ta* (*etxe-eta-n* 'in the houses', but *zenbait etxe-ta-n* 'in some houses'; in *zenbait egun(-e)-ta-n* 'in several days', the *-e-* in parentheses is epenthetic) (cf. §6.3.2 for further details).

The origin of this *-eta* is controversial (cf. §6.3.2). Since Schuchardt, it has often been assumed that it is borrowed from the derivational Latin suffix *-ēta* (as in *fag-ēt-a*, *plurale tantum* of *fag-ēt-um* 'beech grove'), productive for collective meanings such as those involving tree groves (*EMH*: 116-118). But it may also be compared to other elements rooted in the grammar, such as to the derivational suffix which forms verbal nouns in the lexicon (*lapurr-eta* 'theft'), or to the suffix which makes imperfectives from non-finite verbs in the west (Laz. *kanta-eta-n ebe-la* 'as he sang'), and even perhaps to the conjunction *eta* 'and'. This would weaken the Romance origin hypothesis, since it is extremely uncommon to borrow conjunctions, or —as it would be in this case— to create conjunctions from borrowed (derivational) elements. Even so, the question remains open. The fact is that *-eta* appears on demonstratives in the local cases of both the sg. and pl. (see Table 7.1). By contrast, in the noun declension *-eta* appears in the local cases of the pl., but not in those of the sg.

As for the instr., the oldest form on demonstratives is *-(e)z* or *-(e)zaz* (*EMH*: 101-103). Thus: Dch. *hon-ez* 'of this', *horr-ezaz* 'of that', *har-zaz* 'of that yonder'; Lç. *(mundu) hun-ez* 'of this world', *hun-ezaz* 'of this', *(hire ezagutze) horr-ezaz* 'of that knowing you (sg.)'. Later continental authors combine forms such as these with *-taz*, as in Tt. *hon-taz*, E.Cib. *hun-etaz*, *hor-taz*, *hau-taz* 'of these', *he-taz* 'of those yonder'. In the following century, at least E.Sar. only uses *-taz* (*hun-etaz*, *hor-taz* ...) (cf. Camino 2008: 137). It seems that this *-ta-* has been borrowed from the local cases (ines. *hon-etan* ~ instr. *hon-ezaz* → instr. *hon-etaz*, etc.), even if the instr. was not one. In the west, old authors also use *-(e)z* or *-(e)zaz*, as in: Och.A. *on-ez* 'of this', *orr-ez* 'of that', Lar. *arr-ezaz* 'of that yonder', Cb. *orr-ezaz* 'of that', *ori-ezaz* 'of those', Mog. *on-ezaz*, *on-ez* 'of this'. But here, rather than there being a conversion of *-(e)z(az)* into *-(e)taz*, Azkarate & Altuna (*EMH*: 102) suggest that there is a decrease in the instr. in such cases, in favor of periphrases, such as *asko gogora-tzen naiz hare-kin* 'I am often reminded of him/her (lit. with him/her)' instead of *asko gogoratzen naiz har-zaz* 'I am often reminded of him/her'. The *hon-etaz* type, then, which is the instr. form that has prevailed in the standard language, appears to have its origin in the continental Basque Country.

In the plural, the main difference between the case forms of demonstratives and nouns is that, on demonstratives, we find the endings abs. pl. *-ek* and erg. pl.

-ek (see Tables 7.1 and 7.2, columns a, b), whereas on nouns we find abs. pl. *-ak*, erg. pl. *-ak* in the west, but abs. pl. *-ak*, erg. pl. *-ek* in the east. All dialects point to a stage with abs.-erg. pl. *-ek* on demonstratives. Hence, it is probably an old feature, although due to the phonetic developments described in §7.2.2 it has faded in some varieties. In Biscayan —a dialect in which the demonstrative pl. forms have a marked accent on the root—, this *-ek* appears with a geminated vowel in the 1st and 2nd grades from the oldest texts (Cap. Amill. *oneek*, Amill. *orreek*). In some contemporary varieties, this *-eek* has weakened into *-eik*, as in *ón-e(i)k*, *órr-e(i)k* (Ugao-Miravalles), and *on-eik*, *orr-eik* (Aramaio). However, whether it is a reduction of *-eek* or an archaism, *-ek* also appears early (Bet. Mic. VJ. *onek*, although Mic. also has *oneik*, and the text of VJ also presents one case of *oneek*), and the fact that the radical accent (*ón-*, *órr-*) is a pl. marker by itself makes the second *-e-* unnecesary, so that there is no trace of it in modern varieties such as those of Gatika, Lekeitio, or Markina.

This difference between the abs. and erg. pl. forms of demonstratives and nouns is striking. As I see it, this thorny question is as yet unresolved. I shall merely point out that several eastern varieties and authors from different periods have morphologized an abs. pl. / erg. pl. opposition. Thus, Oih. has abs. *hau-k* / erg. *hau-ek* (1st grade), abs. *hori-k* / erg. *hori-ek* (2nd grade), and abs. *he-k* / erg. *hei-ek* (3rd grade) (*EMH*: 16). In the first two cases, the opposition probably arose because, at first, a syncope of *-e-* (*hau-ek* > *hau-k*, *hori-ek* > *hori-k*) was optional or facultative, but was eventually specialized to mark the abs. as opposed to the erg. As for the functional distinction abs. *he-k* / erg. *hei-ek*, this probably arose by means of the specialization of two free variants, such as those found e.g. in Dch.[11] Another Souletin author, Belapeyre, has *hoiek* for both abs. and erg. (1st grade, no opposition), abs. *hori-k* / erg. *hori-ek* (2nd grade, like Oih.), but abs. *hur-ak* / erg. *hai-ek* (3rd grade) (*EMH*: 19). Abs. *hur-ak* came about by the addition of the abs. pl. noun declension ending *-ak* to the 3rd grade demonstrative root *hura*, itself of secondary origin (see §7.2.5). This way of creating a 3rd grade abs. pl. is typical in Souletin, Roncalese and Salazarese, as in Sal. abs. *kur-ak* / erg. *kai-ek* (Arana 2006: 45). Another author who shows such a distinction, but only for the 2nd grade, is Leiçarraga. This has *hau-k* for both abs. and erg. pl. in the 1st grade and *he-k* for both abs. and erg. pl. in the 3rd grade. In the 2nd grade, however, he has abs. pl. *hori-ak*, but erg. pl. *hori-ek* (apud Irigoyen 1987 [1978]: 108). Such an isolated and asystematic change appears to have taken place by analogy with the noun declension.

11 Actually, the three free variants in Dch. are *he-k*, *hai-ek* and *hei-ek*. The latter appears only once, and is probably an assimilation of *hai-ek*.

As for the remaining grammatical cases in the pl., I shall first analyze Biscayan and then the other dialects. B. gen. pl. *ón-een* (1st grade) and *órr-een* (2nd grade) seem to have been formed from the corresponding sg. *on-en* and *orr-en* by doubling the *-e-*, as in the abs.-erg. In fact, the Old B. gen. pl. *-aen*, which appears in the noun declension (Cap. *il-aen* 'of the dead') never appears on demonstratives. Therefore, it appears that *-een* developed within the paradigm of demonstratives (cf. §7.2.6). This *-een* has later undergone by and large the same reduction processes as the *-eek* of abs.-erg. (see above). In the 3rd grade, the gen. has also developed according to the abs.-erg. The old form seems to be gen. pl. *ae(e)n* (Cap. *aen*, *aeen*, Laz. *aen*), but just as *ár-eek* arose for the abs.-erg., gen. *ár-een* and several developments of this form also appear in later authors (as in Añ., Zav., F.Bart., etc.).

The dat. pl. of 1st and 2nd grade demonstratives requires special attention, for here the Biscayan case marker is the independent dat. pl. form *-eri*, which, at least at first sight, cannot be derived from the corresponding sg. of demonstratives, nor from the pl. of the noun declension (west. *-ai*, east. *-ei*).[12] Perhaps it can be compared, however, to S. LN. R. Sal. dat. pl. *-ér* (S. *gizun-ér* 'to the men', Dch. *haier* 'to them', etc.). Azkue (*MV*: 328-329) and Trask (*HB*: 201) derive it from **-eri*.[13] In this case, a dat. pl. ending **-eri*, would have survived in an apocopated form in far east. *-ér* (nouns and demonstratives), and with no apocope in Biscayan (only on demonstratives). Of course, the two forms might also have arisen independently of each other (cf. §7.2.6). The other dat. pl. endings, *-ai* ~ *-ei*, might come from another dat. pl. form, perhaps **-a-ga-i*, although this would require an assimilation *-ai* > *-ei* to explain the latter. Alternative attempts to relate all dat. pl. forms are also conceivable (cf. §6.4.1.3).

Be it as it may, the B. dat. pl. forms of these demonstratives show some diversity. If we start from B. *ón-eri* and *órr-eri* (the forms still used in Gatika), some varieties have dropped the rhotic (as in Aramaio and Markina; already found in Cap. *on-ei* 'to these'), or shifted the accent to the following syllable (as in Lekeitio *on-éri*, *orr-éri*). Another phenomenon, which is apparently an inno-

12 The 3rd grade appears to be different in this respect, since at least in Laz. and Cap. we find dat. pl. *ai*, perhaps from *a-* plus *-i* or *-e-i*, with the dropping out of *-e-*. If this is so, the later *ár-eri* has spread perhaps by analogy with *ón-eri* and *órr-eri*.

13 In fact, dat. pl. *-eri* (adib. *gizon-eri* 'to the men') is attested, and has been spreading throughout the last century, although its attestations are later than those of *-ér* and *-ei*. Pikabea (1993: 47) dates the spreading of *-eri* among Navarrese-Labourdin authors to the 19th century. There are some previous attestations of *-eri*, as in K&B (here we have *-ei* in 14% of the cases, *-er* in 14%, and *-erri* (!) in 72%; see Padilla-Moyano 2011: 42). But this *-eri* is unattested in the earliest texts. So even if B. *ón-eri* and S. *gizun-ér* came from *-eri*, there seems to be no continuity between this and later forms such as *gizon-eri*.

vation, is the extension of dat. pl. *-eri* to the dat. sg., possibly due to the fact that, in the rest of the cases (except the abs.), sg. and pl. forms are morphologically identical, the differentiating characteristic being that pl. forms have a marked accent on the root. Thus, we find *on-erí* 'to this' and *orr-erí* 'to that' in Gatika, and *on-eri* and *orr-eri* in Lekeitio, but *on-i* and *orr-i* in Aramaio and Markina.[14]

In the rest of the dialects, the gen. pl. and dat. pl. forms point to *-en* and *-ei*, respectively. If they were only to be found in the eastern dialects, they might actually be the endings of the noun declension, but it so happens that in the varieties with dat. pl. *-ai*, as in the Guipuscoan dialect, this explanation does not exactly fit for the data. For example, Lar. (1984 [1729]: 8, 26-27) has *Jaún-ai* 'to the men', but *ói-eï* 'to these', *orí-ei* 'to those'. Perhaps the dat. pl. of demonstratives, then, arose within the paradigm of demonstratives, the *-e-* being extended from the abs.-erg. *hau-e-k* and *hori-e-k* to the gen. and dat. Just as in those forms we have a structure /stem + *-e-* (interpreted as a plural marker) + case marker/, if this structure spread to the dat. pl. we would have *hau-e-i* and *hori-e-i*.[15]

In the local cases, we have to start from *ón-/órr-* + *-e-* + *-eta-* + case marker in Biscayan (Cap. *on-eetan* 'in these', *orr-eetarean* 'from these', VJ. *on-eetarik*), *hau-/hori-* + *-eta-* + case marker to the east of Biscay (and in Laz.). As in the sg., the *-e-* of *-eta* may drop in the 1st grade in some continental authors, as for example in Ax. Oih. *hau-tan*, K&B. *ho-tan*, but since demonstratives are definite, we can presume that this is a result of phonetic loss. In animates, the oldest pattern for forming the local cases appears to be, in the 3rd grade, *-gan-* upon *har-* (Tt. *har-ganat*, Dch. *har-gana* 'to him', Laz. Cap. *a-gan* 'in him'). Lç. and the Old L. authors, however, attach *-gan-* to the genitive (Lç. *haren-*

14 Already in Capanaga I find two cases of *on-eri* 'honi' (no *on-i*), and also Cap. *albo odolez bete orr-eri* 'to that side full of blood', but also *zuri aita espiritual-orr-i* 'to Thee spiritual Father'. We also have RS. (422) *arz orr-i* 'to that bear'. Similarly, in the *Chant of Olaso* (traditional, first recorded by Gar. in the 16th century, referring to events which took place in the mid-15th century), we find *Berba orr-i* 'to that word' (see *TAV*: 88-90), but in *Abendaño's Lamentation* (also traditional, going back to the mid-15th century) we find *aldi-oneri* 'to this awkward situation' (see *TAV*: 79-81). Urgell (2001: XC) also considers *-eri* on *on-*, *orr-* and *bat-* 'one, a' to be an innovation (in the case of *bat* it is very clear; see *OEH*: 'bat'). In Añ.'s *Esculiburua* of 1802, he uses only two examples of *-eri*, but in the 1821 edition, the number of forms like *on-eri*, *orr-eri*, *bat-eri*, has increased. If we look in the *CA*, we can see that in the first half of the 18th century, Barr. exclusively uses *on-i* and *orr-i*, and in the second half of the same century, the same holds for Mog. (only one case of *on-eri*) and Ol.

15 Cf. §7.2.6. It should be pointed out that Old L. authors like E.Cib., Ax., Pouv., Gç., Urt. and Haran. use *hau-i* for *hau-ei*. This appears to be the same syncopating process as that of (e.g. Ax.) *hau-ek* > *hau-k*, *hau-etan* > *hau-tan*, unless we believe that there has never been any *-e-* there.

gana, E.Cib. *haren gana* 'to him'). The possibility of attaching *-gan-* to the old form of the gen. is also attested, as in Mst. *hare-gana* (cf. Camino 2008: 140 for details on continental texts). In the 1st and 2nd grades, *-gan-* was usually attached to the recent form of the gen. in the continent, although we find at least *hone-gana* 'to this', upon the old form, in Dch. This possibility, however, has been more successful in the west (cf. *onegan-* vs. *onengan-*, *orregan-* vs. *orrengan-*, even *aregan-* vs. *arengan-*, in the *CA*).

A final complication to the picture outlined in §7.2.3 is the fact that, in the lower course of the Bidasoa River and in some adjacent areas (corresponding roughly to the northern part of the North.-HN. dialect and the southern part of the L. dialect), in the 1st and 3rd grade demonstratives, all or some grammatical cases other than abs.-erg. and/or all or some local cases are formed by suffixing the case markers to the abs.-erg. form. For the contemporary Basque of the Baztan Valley (to the east of the low Bidasoa area), some of the forms provided by Salaburu & Lakar (2005: 68-69) are in Table 7.3.

Table 7.3. Some plural forms of demonstratives in the Baztan Valley

	1st grade	**2nd grade**	**3rd grade**
Abs.-Erg. pl.	*hok*	*hoiek*	*hek*
Dat. pl.	*hoki*	*hoiei*	*heki*
Gen. pl.	*hoken*	*hoien*	*heken*
All. pl.	*hoketara*	*hoietara*	*hetara*
Ines. pl.	*hoketan*	*hoietan*	*hetan*
[Salaburu & Lakar (2005: 68-69); a selection]			

Note that, in these forms, this procedure of forming the plural has not reached the local cases of 3rd grade demonstratives, and has left the 2nd grade column untouched. In fact, if we compare these forms to those of Ax. (born in Urdazubi, to the north of the Baztan Valley), we can see that this class of plural formation is present in some forms (1st grade gen. *hauk-en*, 3rd grade gen. *hek-en*, compare e.g. Lç. *hau-en*, *hai-en*), but not in others which are present in Table 7.3 (1st grade dat. *hau-i*, ines. *hau-tan*, 3rd grade dat. *hei*). In Ultzama, to the south, analogical forms have not reached the 3rd grade either (Ibarra 1995: 328). I do not think, then, that these forms are anything other than innovations.[16]

16 In any case, as Apalauza (2008: 94-95) notes, analogical forms like these are known to be found further to the south, such as in the Imotz valley.

7.2.4. The demonstrative → article phenomena

The Basque definite article comes from the cliticization of the 3rd grade demonstrative, as is also the case in the neighboring Romance languages (e.g. Lat. *illa mulier* 'that woman yonder' > Sp. *la mujer* 'the woman'; cf. §6.3.1). However, in Basque the demonstrative was postponed, so that the article was suffixed. On the other hand, the article has further penetrated into many more contexts than in Romance, for instance into subject complements (*neska ederr-a da* 'the girl is pretty (lit. the-pretty)'). This occurred because *-a* was exploited to mark singulative meaning in a language which lacked number (i.e., the morphologized opposition between sg. and pl., and/or other number notions) as a grammaticalized category (see Irigoyen 1987 [1985b]), and §6.3.1).

This conversion of the 3rd grade sg. demonstrative into the sg. article did not cause this demonstrative to disappear. Thus, the opposition between *gizon-ari* 'to the man' (< *gizon har-i*) and *gizon hari* 'to that man yonder' is, at least in SB, an accentual one (the former is treated as one accentual unit, the latter as two), reinforced by the aspiration in those dialects which preserve it (or by *k-* / *g-* in peninsular eastern varieties). In the abs. sg. alone, to make up for the loss of **har* 'that yonder', all dialects except Biscayan have developed the form *hura* 'that yonder' (east. *kura*, *gura*). According to Irigoyen (1981: 391-394), the origin of this *hura* would be *haur- + -(h)a(r-)*, that is, the ancient form **har* reinforced on the left by *haur-* in its original meaning of 'ecce' (see §7.2.5). This **haur-har* would usually develop into *(h)ura*, but more archaic phases of that development such as *aura* survive, according to Irigoyen, at different points of the Deba Valley (also *auré* in Azpeitia and *aré* in Azkoitia), and *ore* can be found in 18th and 19th century texts from Goñi and Oltza (South.-HN, several kilometers to the west of Pamplona-Iruñea).[17]

However, unlike Romance, Basque also shows traces of attempts to turn the 1st and 2nd grade sg. demonstratives into articles. Remaining examples of the 1st grade sg. demonstrative conceived as an article can be found in medieval person names such as *Lope Hederr-o* vs. *Garcia Hederr-a* (the former being something like 'Lope, you that handsome one', and the latter simply 'Garcia the handsome'; Navarre, 14th century), and also in Old L and Lç., e.g. Lç. *seme gazten-or* (< *-(h)aur*) 'this younger son'.[18] This can also be found in western dialects,

17 Camino (2004: 86) points out that *ore* is found in Goñi, Olloibar, Arakil, Oltza, Txulapain, Atetz, Odieta, Ezkabarte, Ultzama and Anue, and *ora* in Goizueta. The data from the Deba Valley are more controversial as evidence in favor of **haur-har*, for *aura* and related forms are very poorly attested before 19th century authors (only three cases in Peñ. and two in Ag.; see *OEH*: 'hura' and *CA*).

18 See *EMH*: 41, 27.

where the cliticization of *-au* (< *-(h)aur*) is often represented accentually (in speech, as one accentual unit) or graphically (in texts, as one joined word), as attached to the modified NP, as in B. G. *gizon-au* 'this man here' or B. G. *ni bekatari-au* 'me sinner', *ni-ri bekatari-oni* 'to me sinner', and so on.

This latter example shows a commonplace correspondence between the 1st grade cliticized deixis and the 1st person in the west. This correspondence does not appear in the example *Lope Hederr-o* 'Lope, you that handsome one' above, where *-o* (< *-hau(r)*) indicates 2nd person. Perhaps this is the case because *haur* originally indicated proximate deixis, which could include, apart from a proximate 3rd person, the 1st or 2nd person (see §7.2.5). Once *hori* (> *-ori*) arose, it was usually associated to the 2nd person (Cap. *zu-ri Aita espiritual-orri* 'to you spiritual Father', Laz. *zeure buru-ori* 'that head of yours', Dch. *larrui-ori oso du-zu* 'that skin (of yours) is healthy', Lç. *hi bekhatu-tan sorthu-a aiz guzi-ori* '(lit.) the whole of you is born in sin'), and this restricted *hau* to the 1st person. In some other cases, *-ori* expresses a different kind of proximity, corresponding to the 3rd person: Laz. *Silbia, dama galant-ori / pena handi-az bizi zan* 'Silvia, that gallant lady, lived with a great pain', Dch. *gorpuz-ori ustel-tze-ra lur hotz-ian egoiz-ten* 'that body is thrown to the cold earth to rot away'.

In the plural, the 1st grade demonstrative *hau-ek* also appears cliticized, but evolved into abs.-erg. *-ok* (dat. *-oi*, gen. *-on*, ines. *-otan*, etc.) in the oldest texts of every dialect. As a 1st grade demonstrative, it is usually associated to the 1st person (Cap. *gu bekatari-ok* 'we sinners', Dch. *bi-ok* 'the two of us', Dch. *guzi-or* 'to all of us'), but it can also appear related to the 2nd person (Laz. *zeure begi-okaz* 'with those eyes of yours', Dch. *zure pena dio-zu-n-ok* 'those sorrows of yours that you mention'), or to a proximate 3rd person (Cap. *Berba oneek eur-ok* 'These words themselves'). At the beginning of the written tradition, it was already on the wane in eastern dialects (see, however, Camino 1997: 368-370), but in Biscayan and Guipuscoan (and Alavese) it has consolidated and is currently a proximate article in complementary distribution with the article *-ak* (*-ai*, *-en*, etc.).

There seems to be, then, a tendency to relate deixis to person. This process appears to embrace two stages. At the first stage, *haur* (> west. *-au*, Old L. *-or*) is associated with 1st and 2nd persons, or with the 3rd person, but here implying some kind of proximity, and at the second stage, *hori* (> *-ori*) breaks into the deictic system, typically associated with the 2nd person but sometimes also with a proximate 3rd person. This relationship between deixis and person will become clearer when we discuss personal pronouns (see §7.3.3).

In Biscayan, there is a pleonastic demonstrative construction consisting of 'demonstrative + NP + suffixed article / demonstrative'. Table 7.4 shows examples from Lekeitio (only abs. and erg. forms).

Table 7.4. NP-initial demonstratives in Lekeitio (abs., erg.), exemplified by ume *'child'*

a.		'this child'	'these children'
	abs.	*au umi-a ~ au umi-au*	*ónek umí-ak/umí-ok*
	erg	*onek umi-ak ~ onek umi-onek*	*ónek umí-ak/umí-ok*
b.		'that child'	'those children'
	abs.	*ori umi-a ~ ori umi-ori*	*órrek umí-ak/umí-ok*
	erg.	*orrek umi-ak ~ orrek umi-orrek*	*órrek umí-ak/umí-ok*
c.		'that child yonder'	'those children yonder'
	abs.	*a umi-a*	*árek umí-ak*
	erg.	*arek umi-ak*	*árek umí-ak*
[Hualde & Elordieta, G. & Elordieta, A. (1994: 98-99); a selection]			

Constructions like these appear from the beginning of the written tradition in Biscayan (e.g., VJ. *oneek amar Mandamientu-ok* 'these ten Commandments'; apud Irigoyen 1981: 367). In spite of this, rather than being an archaic feature, such constructions appear to be a means of creating new demonstratives, once the original suffixed demonstratives have lost some of their deictic force, when they are cliticized to the NP modified by them (a tendency particularly strong in western dialects). Hence, they may be an innovation.[19]

7.2.5. The internal structure of deictic roots and the origin of the tripartite deixis

The three-deixis system depicted in Table 7.1 exists in all dialects, so that it must be postulated for the common language, but probably only for the sg. (see §7.2.6 for the pl.). In the sg., the dimorphic root structure is apparently common: abs. *hau(r)* / *hon-* in the 1st grade in the remaining cases, abs. *hori / horr-* in the 2nd grade. Even the 3rd grade must be common, pointing to abs. **har* / *har-* + case marker in the rest of the cases, in spite of the existence of abs. *hura* outside Biscayan,[20] and other innovations such as e.g. erg. *(h)ar-e-k* in different varieties.

Where does this three-deixis system come from? According to Irigoyen (1981), it derives from a previous two-deixis system, which, due to the common character of the tripartite system, must be pre-common (i.e. Proto-Basque). In fact, Irigoyen's reasoning is an exercise of internal reconstruction which starts

19 Unless we relate them to similar eastern constructions found e.g. in Aescoa (Camino 1997: 414, 2004: 77), such as *Gura mendi ura, biziki mendi on-a da* 'That mountain over there is a very good mountain'. I still think that these are innovations, although their similarity makes a common origin worth considering.

20 In fact, the B. 3rd grade abs. demonstrative, as in *a etorri da* 'that yonder has come' appears to be a remnant of abs. **har*.

from the bipartite system. In Irigoyen's account, there were originally two deictic elements: the proximate *haur* and the distal **har*. The former would originally be a kind of interjectional adverb with a meaning close to that of Lat. *ecce* 'behold (here/there)!'. Remnants of this adverbial meaning seem to be present in attestations of far eastern authors, as in the following proverb compiled by Oihenart:[21] *haur ola zirola*. This was not understood by the first editor, but Irigoyen assigns to it the literary meaning 'Behold there your workshop, shoemaker!'. The literal meaning would be 'Right here/there your workshop, shoemaker!'. The adverbial meaning would also be found in the place adverb of 2nd grade *hor* 'there', common to all dialects, which is apparently a phonetic evolution of the same *haur* of Oih.'s proverb.[22] This 'ecce' sense of *haur* also appears in Oih.'s poems, as in *haur non datza-n ehortzi-a* 'here is where he lies buried', with *haur* prompting the complementizing suffix *-n* in the finite form, or even with *haur* inflected as abs. pl. in *Hori erran-ik, hau-k doaza-la ...* 'That said, behold, off they go ...'.

In the process of demonstrative formation, however, this *haur* seems to have split into two grades, 1st and 2nd, opposed to the 3rd grade **har*. The abs. of the 1st grade is simply *hau(r)*; the abs. of the 2nd grade is *hor-i*, according to Irigoyen < **haur-i*, perhaps *haur* + the dative case.[23] Figure 7.1 represents this hypothesized development.

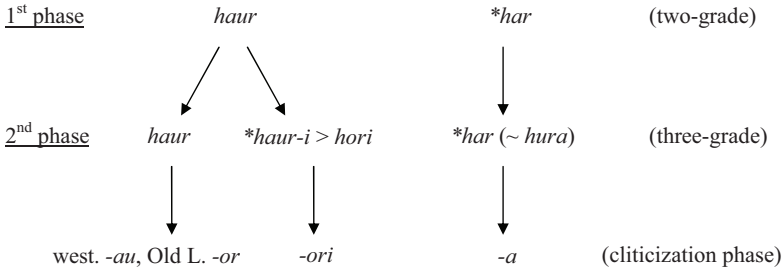
Perhaps the relationship between deixis and person mentioned above (see §7.2.4), and especially the intensifying processes undergone by 1st and 2nd p. pronouns by means of adverb/demonstrative suffixing (see §7.3.3) can support this view. By the first of these processes, both 1st and 2nd p. pronouns are reinforced by *haur*. By the second, *haur* is restricted to 1st p. pronouns, while *hori* takes over 2nd p. pronouns. This points to a secondary rise of *hori* as an intermediate grade.

---

21 It was not published in 1657, but was kept in a manuscript and published in 1967 (Irigoyen 1981: 376). It was collected in Soule.

22 The three place adverbs are: 1st grade *hemen* 'here', 2nd grade *hor* 'there' and 3rd grade *han* 'over there'. The 3rd grade one would be the distant element *ha-* plus the inessive *-n*. The 1st grade *hemen* was derived by Michelena (*FHV*) from the 1st grade demonstrative: **hau-en* '~ in this' > **haw-en* > **hab-en* > *(h)eben* (preserved in Alavese and eastern dialects).

23 This is the assumption made by Irigoyen (1987 [1978]: 108), who puts *hor-i* in relation to some other (much more transparent) datives turned into absolutives such as *ner-oni* 'I myself', etc. (see below in §7.3.3). However, since *hori* (> *-ori*) is closely related to the 2nd person when it enters the deictic system, and actually snatches this semantic realm from *haur* (see §7.2.4 and §7.3.3), an alternative possibility is that **haur-i* is simply *haur* 'this' + *(h)i* 'you (sg., familiar)'.



[*EMH*: 42, somewhat modified]

Figure 7.1. Irigoyen's (1981) proposal for the rise of the three-deixis system

In any case, *hau(r)* and *hori* are the roots for the abs. sg. All the other cases in the sg. are built upon a different root: 1st grade *hon-*, 2nd grade *horr-* (see Table 7.1). But, where do these come from? Irigoyen (1981: 373) derives *hon-* from **hau-n-*. Thus, we would have abs. sg. *hau-r* as in origin, but **hau-n-* in the rest of the sg. cases. If so, this alternation may be compared to a series of alternations in the patrimonial lexicon of Basque whereby *n* and *r* alternate, as in *egun* 'day' / *egu-r-aldi* 'weather' and *jaun* 'lord' / *jau-r-egi* 'palace'. A drawback of this comparison is that the alternation found in the 1st grade demonstrative *hau-r* / **hau-n-* is the opposite of the one found in the lexicon, where we have *-n* in word-final position and *-r-* in non-final position.

In the 2nd grade, the abs. sg. is *hor-i* and the other cases are based on the root *horr-*. The final *-i*, then, is exclusive of the abs. As for the multiple rhotic of the rest of the cases, this may be a generalization of the *-r* of *hor-*, allophonically uttered as strong in some phonetic contexts. Nevertheless, such a context should logically be one before a stop, but erg. **hor-k* or a similar form is unattested.[24]

24 I take this possibility from Irigoyen (1981: 383). Although he refers to the 3rd grade, it seems to me that he has in mind the same development for the 2nd grade. In the 3rd grade, we have erg. sg. *har-k* (currently the most widespread form), whose rhotic uttered as multiple (vs. simple gen. *har-en*, dat. *har-i*, etc.) has spread to HN. gen. *harr-en*, G. *arrez-kero* ('after that', built upon the instr.), and so on (see *FHV*: 331). Although these cases are late and dialectal, if we assume a similar process for the 2nd grade at a common period, the explanation becomes conceivable —but obviously hypothetical— that, for instance, an erg. **hor-k* 'that' was allophonically uttered as multiple and that this multiple spread to all of the cases other than the abs., even if the form which eventually prevailed as erg. sg. had an epenthetic vowel (common erg. *horr-e-k*).

To sum up, all dialects have a three-deixis system of demonstratives, whose abs. sg. forms point to *haur*, *hor-i* and **har*, respectively. In the first two, all the cases other than abs. are built upon *hon-* and *horr-*. All of this must already be Common Basque. If we resort to internal reconstruction, an older stage can be hinted at, apparently predating the demonstrative formation proper, where there were two grades: a proximate *haur* —perhaps an adverb in origin— and a distal **har*.[25] From this situation, an intermediate grade arose from the 1st grade. According to Irigoyen's reasoning, we would have a dimorphic *haur* / **hau-n-*. Thus, *haur* —'ecce' in origin— became the abs. sg. of the 1st grade, and the rest of the cases arose from the monophthongization of **hau-n-* > *hon-*. Meanwhile, the monophthongization of *haur* > *hor-* gave rise to the 2nd grade, as both place adverb (*hor* 'there') and demonstrative (*hor-i-*, with a suffixed *-i*). The rest of the cases were built upon the root *horr-*, with no *-i* but with multiple rhotic, which perhaps spread from a context where it was allophonic.

7.2.6. The category of number on demonstratives

The singular paradigm of demonstratives, then, had already been built by the period of dialectal fragmentation.[26] What about the plural? Here we have two different questions: the root and the endings. The root of the plural of 1st and 2nd grade demonstratives, to which I shall limit my analysis, shows an interesting dialectal split (cf. Table 7.2): B. *hón-*, *hórr-* / *hau-*, *hori-* in the rest of the dialects. As for the endings, if we leave aside some dialectal differences in some cases (e.g. the dat., see §7.2.3), in general they are also common. What is most striking is the difference between the abs. and erg. case endings of demonstratives (pointing to abs. *-ek* and erg. *-ek* in all dialects) as compared to the same endings in the noun morphology (pointing to abs. *-ak* in all dialects, and to west. *-ak*, east. *-ek* for the erg.). This mismatch has never been satisfactorily explained.

---

25 Lakarra (2013, §5.7, among others) has proposed *har(-tu)* 'to take' (attested as such in all historical dialects) as the lexical source of this **har*. Such a development has good typological parallels (cf. Lord 1993). As for *haur*, Ariztimuño (2011) has pointed to the word *ahur* 'palm (of the hand)' of continental dialects as a good candidate to be its lexical source. Another related word might be west. *aurre* 'front part', with paragogic *-e-* (< *aurr-e-*).

26 Obviously, I am referring only to the structural form of demonstratives. Case markers suffixed to demonstratives show the same dialectal distribution which they have in the noun declension (e.g. B. *horr-e-gaz* 'with that' / *horr-e-kin*; cf. §6.5.2.1).

I will not attempt to do this myself. But, wherever the abs. pl. and erg. pl. forms in *-ek* of demonstratives come from, what is incontrovertible is that the west and the center/east built the plural paradigm upon different roots (cf. Table 7.2). Whereas the west built this paradigm upon the roots which correspond to the cases other than the abs. in the sg. (*hon-*, *horr-*), the center and the east opted for the abs. sg. forms *hau(r)*, *hori*. This points to a dialectal and late formation of the plural of demonstratives (just as in other morphological realms). I would rather like to show that this difference in the choice of the root brought consequences for the morphology of the whole paradigm.

In the west, once abs.-erg. pl. *(h)on-ek*, *(h)orr-ek* —which I assume to be the original forms— were morphologized, and the whole paradigm was built upon *hon-*, *horr-*, the procedure was: /root + case marker/. Except for the abs., this generated homophonous pairs, as the outcomes would be gen. (sg. and pl.) *(h)on-en*, dat. (sg. and pl.) *(h)on-i*, etc. This must be the reason why the western demonstratives have developed two differenciating devices for the plural which are not present in the other dialects: the doubling of the *-e-* and a marked accent on the root.[27] How exactly this occurred is not an easy question. Perhaps the *-e-* of *-e-k* simply spread to the whole paradigm from the abs.-erg. pl. forms once these had been morphologized, because it was reinterpreted as a pluralizer within the demonstrative paradigm. Thus, gen. pl. *(h)on-en* became *(h)ón-e-en*, dat. pl. *(h)on-i* became *(h)ón-e-i* (perhaps with a later replacement of *-i* by the new dat. form *-ri* created by reanalysis, cf. §6.4.1.2, §6.4.1.3), and similarly in the local cases: ines. pl. *(h)ón-e-etan*, etc. To sum up, according to this view the *-e-* of *(h)on-e-k* was interpreted as a pluralizer and spread throughout the plural. This process also recharacterized the plurality of the abs.-erg. pl. itself (hence → abs.-erg. pl. *(h)on-e-e-k*).

In the center and east, the plural paradigm of demonstratives was built upon the roots corresponding to the abs. in the sg.: *hau(r)-*, *hori-*. For whatever reason, though, the abs.-erg. pl. endings were —unlike the root— apparently the same as in the west, i.e. *-ek*. Be it as it may, here the sg. and the pl. were built upon different roots, which rendered differenciating devices unnecessary (erg. sg. *hon-ek* / erg. pl. *hau-ek*, gen. sg. *hon-en* / gen. pl. *hau-en*, ines. sg. *hon-etan* / ines. pl. *hau-etan*), etc.[28]

27 But it is difficult to determine in which direction this accentual distinction between e.g. sg. *on-e-k* / pl. *ón-e-e-k* may have developed, i.e. whether the sg. form lost a marked accent or the pl. form developed a marked accent (see, on this point, Martínez-Areta 2004: 201).

28 It cannot be ruled out, however, that, at least in the dat. pl., a spread of the *-e-* from the abs.-erg. pl. *-ek* recharacterized the plurality of *hau-i* (which is actually attested, cf. §7.2.3) and **hori-i* (?), thus yielding *hau-e-i*, *hori-e-i*.

As I said, rather than the chronological prevalence of one system over the other, this diversity in the formation of the plural might point to a state of the grammaticalization of the plural which was still *in fieri* during the dialectal period, so that for this process the Biscayan dialect and the rest of the dialects took different paths.

7.3. Personal pronouns

7.3.1. 1st and 2nd person pronouns

Historical Basque has the 1st and 2nd p. weak (or neutral) pronouns shown in Table 7.5.

Table 7.5. 1st and 2nd p. pronouns in historical Basque

	sg.	sg. (< pl.)	pl.
1st person	*ni* 'I'		*gu* 'we'
2nd person	*hi* 'you (sg., familiar)'	*zu* 'you (sg., formal)'	*zuek* 'you (pl.)'

The familiar treatment (*hi-ka* in Basque grammatical terminology) requires allocutive finite forms of the verb (a morpheme must represent the addressee even when (s)he is not an argument of the clause; see §9.6). The use of this treatment, however, has been historically on the wane, and a number of varieties (such as Ondarroa, B) no longer use it, so that the formal treatment (or *zu-ka*) is the only one used. In the varieties where *hika* is used, the scope of *hi* is more restricted than the French *tutoiement*, and much more so than the Spanish *tuteo*. Azkue (*MV*: 433) defines the *hi-ka* treatment as that of comradeship. Alberdi (1995: 289) points out that, within the family, "asymmetric treatment *hi/zu* according to age relations is common (parent/child, uncle or aunt/nephew or niece, grandparent/grandchild, etc.)", whereas "[o]utside of the family, […] there is a tendency to replace asymmetric treatment with the same treatment in both directions". The same author further remarks that the scope of *hi* has not significantly changed since the 16th century. As is well-known, however, in Leiçarraga's works God is referred to with *hi*, which could point to a wider scope in previous centuries.

Some peninsular varieties of B, G and HN have developed a third 2nd p. pronoun, more formal than *zu*: *ber-ori*. It is formed by attaching the 2nd grade demonstrative to the element *ber-*.[29] It requires 3rd p. agreement on the corre-

29 In fact, the simple form *ori* for *berori* is also used in several Navarrese valleys, as Azkue (*MV*: 432) reports.

sponding finite verb, so it looks like an imitation of the Sp. *usted* (< *vuestra merced*). Similarly, in several far eastern areas (such as Sal and LN of Cize) another ternary opposition is possible, because the pronoun *xu*, intermediate between *hi* and *zu* in the formality scale, has arisen. This is the palatalized —hence, more familiar— variant of *zu*.

From the point of view of the prehistory of the language, it is very interesting that *zu* must originally have been a 2nd p. pl. On purely internal grounds, this pronoun triggers the presence of the same pl. morpheme as the 3rd p. pl. (*da-bil* '(s/he) walks', but *da-bil-tza* '(they) walk', *za-bil-tza* '(you (sg.), formal) walk'). We should further note that *ni* and *hi* have the same vowel, whereas that of *zu* is identical to that of *gu*. The significance of this vowel alignment seems to be related to the number category (sg. *i* / pl. *u*). On the other hand, even if from the oldest texts we have *zu* only as a sg. form, in a number of western attestations the corresponding agreement markers *-zu*, *za-*, *ze-* are still plural (e.g. Laz. *Zu-ek jakin-go do-zu* 'you (pl.) will know').[30] There seems to have occurred, then, a process similar to that of several European languages (e.g. Spanish, French, and English), whereby a 2nd p. sg. form of respect has been created from the 2nd p. pl. Consequently, the language has developed another 2nd p. pl. by attaching the pluralizer *-ek* for both abs. and erg., possibly under influence from the Spanish model (*vos* 'you (sg.)' → *vos-otros* 'you (pl.)'). Since *zu-ek* is morphologically common, either the process had already taken place in the common language or it affected all dialects in a post-common period.

7.3.2. Peculiarities of the case forms of 1st and 2nd p. weak pronouns

The abs. of the pronouns described in Section §7.3.1 is the bare root. The erg. is formed by simply attaching the erg. *-k* (*ni-k* 'I (erg.)', etc.). However, the 2nd p. pl. has *zu-ek* for both abs. and erg., and some western B. varieties have *zu-ok* (and gen. *-on*, dat. *-oi*, etc.) (Txorierri) or *zu-r-ok* (Baranbio) instead, according to Azkue (*MV*: 431).[31] The *-ek* ending (abs. and erg. in all dialects) must have been borrowed from demonstratives. In fact, it has also penetrated the pronoun *gu*, for we already find erg. *gu-ek* in a Basque sentence in the Emilian Glosses (from as early as the 10th or 11th century!). We also find *gu-ek* as both abs. and

30 Azkue (*MV*: 433) reports that, until his lifetime, in some Biscayan villages such as Bakio and Eibar (in Guipuscoa, but linguistically Biscayan) this archaism (the verbal markers of *zu* with plural meaning) has survived.

31 In fact, *zu-ok* is possible as a variant of *zu-ek* wherever *-ok* exists as a proximate pl. article, in principle.

erg. in Laz. (although he also has the regular abs. *gu* and erg. *gu-k*), and at least abs. *gu-ek* in Mic. (*OEH*: 'gu').

The dat. consists in attaching the pure case form *-i* to the pronoun with epenthesis of a simple rhotic (*ni-r-i* 'to me', *gu-r-i* 'to us', etc.). The origin of this rhotic is controversial. It could come from a reanalysis of the sg. of the noun morphology, whereby *gizon-ar-i* 'to the man' (< **gizon-har-i*) was reinterpreted as *gizon-a-ri* (root-article-dat.), and from there *-ri* jumped to other morphological realms (cf. §6.4.1.2 and §6.4.1.3). Alternatively, it may be derived from the reinforced pronominal forms (abs. *ni-haur* → dat. *ni-haur-i* > *neur-i*, reanalyzed as *neu-ri*, hence → *ni-r-i*), although this process is not common to all dialects. The gen. —and the cases deriving from it, e.g. soc. *-(r)e-kin*, etc.— is formed by the same procedure (and hence its rhotic might have the same explanation), but here the case marker is *-e*, hence *hi-r-e* 'of you (sg.)', *gu-r-e* 'of us', etc. This gen. *-e* is not the same as that of the noun morphology (*gizon-ar-en* 'of the man'), nor of proper nouns (*Aitorr-en* 'of Aitor'), as it lacks *-n*. Instead, *-e* seems to be an archaic form of the gen., also present in some cases derived from the gen. (soc. *gizon-ar-e-kin* 'with the man', probably < **gizon-ar-e-kide-n* 'in the company of the man'; cf. §6.5.2.1), and perhaps even in a couple of medieval attestations such as *Lope iaun Orti-r-e Seme-a* 'Lope, son of Sir Orti' (in the documentation of the Navarrese monastery of Iratxe, 1125; cf. §6.4.1.2). The 2nd p. pl., however, has always been the more recent gen. *-en* (*zu-en* 'of you (pl.)').

However, the expected 1st p. gen. form, *ni-r-e*, does not appear in the oldest texts of any dialect. Instead, the gen. of the weak 1st p. sg. pronoun is *ene*, which is morphologically mysterious. If we start —as I propose— from **nen-e* (later > *ene*),[32] we might conceive that **nen-e* was the old gen. *-e* attached to *nen-*, the index of 1st p. sg. abs. and (sometimes) erg. (actually *ne-*) in finite verb forms of the past and irrealis conjugations. In fact, the presence of a morpheme identical with a verbal index in the pronoun paradigm is otherwise unparalleled, but despite this, *ene* must undoubtedly be brought back to Common Basque.

Be it as it may, *ni-r-e* arises by analogy with *hi-r-e*, *gu-r-e* and *zu-r-e*. According to the *CA*, *nire* —and derived forms like *nire-tzat*, etc.— first appears abundantly in the second half of the 18th century, in Biscayan authors such as Mog. and Ol. (along with some others). *Nire(-)* arises, then, in the west, espe-

32 Such a reconstruction and dissimilation of the first nasal would not be unparalleled in Basque. Lakarra (1995, and some later works) has suggested that in a number of lexical items like *ohol* 'plank' (R. *õl*), reconstructed by Michelena (1988 [1950]: 195-196) as **onol*, we might postulate **no-nol* —the reduplication being a productive device at some older period—, which further dissimilated into **onol* > *o(h)ol*. Some other reconstructions of this type proposed by Lakarra are **na-nan-tz-* > *ahantzi* 'to forget', **ni-nin-tz* > *ihintz* 'dew', and **na-nal* > *ahal* 'can'.

cially in Biscayan, and here *ene* becomes restricted to exclamative expressions and similar constructions. Another form which sweeps *ene* aside in some regions —or which restricts it to marginal constructions— is *nere* (most probably < *neure*). In the *CA*, there are three cases of *nere* in Labourdin authors from the first half of the 17th century, but given the context, they appear to be a variant of *neure*. In the following half there are no attestations, and in the first half of the 18th century there are no fewer than 611, but here nearly all of them are found in Guipuscoan authors (in detriment of *ene*, but tending to become the only form, as was already the case in Mend. and Ub.; cf. Sarasola 1980: 439). *Nere*, then, seems to irradiate from Guipuscoa, but is the standard form in the western areas of Navarre such as Baztan (Salaburu & Lakar 2005: 67), Ultzama (Ibarra 1994: 323), and the Middle Sakana (Erdozia 2001: 179). As we move to the east, *ene* also appears, as in Aescoa (Camino 1997: 406, with data collected by Bonaparte). The same holds with respect to the continental Basque Country. In Bourciez' collection (Aurrekoetxea & Videgain 2005: 28-29), *nere* prevails in 6 municipalities, and *ene* in 132. The six of *nere* are found in the west of Labourd.

In the dat. case, *ni-r-i* is the form attested in old texts, even e.g. in Dch. (LN, 16th) and Tt. (S, 17th c.). From the 18th c. on, the form *eni* breaks through in LN. and S. authors. This appears to have arisen by analogy with gen. *ene*, unless for some reason (a preference for LN. and L. forms?) it remained hidden in old texts but is actually the archaic form (in such a case, perhaps from **nen-i*, parallel to gen. **nen-e*). From the 19th century on the analogy with gen. *nere* seems to have triggered a dat. *neri* in Guipuscoan, from which it may have spread to neighboring dialects (cf. *OEH*: 'ni', for both *eni* and *neri*).

In the plural persons, *gere* and *geren* are found for *geure* in Guipuscoan and High Navarrese authors from the 17th century. In *geren*, *-n* may have spread from *zu-en* 'yours (pl.)', where it indicates plural form. The same explanation can hold for L.Elc. *geuren* and similar variants (see *OEH*: 'geure').[33] Also in Guipuscoan, from the 18th century, we occasionally find *zere* for *zeure*.[34]

The instr. of personal pronouns runs parallel to that of demonstratives (§7.2.3), although here the common form must be exclusively *-(t)zaz*, which perhaps might be analyzed as a pleonastic instr., the first one keeping a form with *-a* (i.e. *-(t)za* + *-z*). This one (e.g. Laz. *ni-zaz akordadu-rik* 'reminded of me', etc.) is the only form in 16th century texts, and also in Labourdin texts of the first half of the 17th century (Ax. E.Cib. *ni-tzaz*). By the 17th century, howev-

33 From the 18th century onwards, a form which has spread and which is considered by Camino (2004: 66) to be an exclusive unitary feature of South.-HN (*Southern Navarrese* in his terminology) is *gauren* 'geure', with a —somewhat unnatural— change *eu* > *au* which is prevalent in this dialect.

34 *Zuek* has too many variants to summarize here (see *OEH*: 'zuek, zeurok, zeuen').

er, in continental authors we already find Tt. *ni-taz*, *zu-taz* (he also has *ni-(t)zaz*, *zu-(t)zaz*), Pouv. Arg. *ni-taz*, *zu-taz*, etc., and the variant with *-ta-* becomes, with few exceptions, the exclusive form in Labourdin, from the 18th century onwards (see *CA*; cf. also Camino 2008: 137, for further details on this point). In peninsular texts, the attestations with *-ta* are much later. As in demonstratives, *-taz* seems to have extended from the *-ta* of local cases of the pl. and indef. of the noun declension, and from the sg. and pl. of demonstratives (ines. *hon-etan* ~ instr. *ni-(t)zaz* → instr. *ni-taz*, etc.).

Finally, since personal pronouns are animate, they form the local cases by means of *-gan*, e.g. *hi* 'you (sg.)', → *hi-gan* 'in you', *hi-gan-a* 'to you', *hi-gan-(d)ik* 'from me', etc. Another possibility is to attach the suffix *-gan-* to the gen.: *hire-gan*, etc. These are the two possibilities found in continental varieties from the beginning of the tradition. In Laz., both possibilities are found, apparently in free variation, although later continental grammarians usually give the type *(h)i-gan* as the only possibility (*OEH*: 'hi'), and in fact this option is much more attested in the tradition (see *CA*).[35] In the continent, there are more possibilities. Apart from those already mentioned, the option exists to form the local cases by means of *-ta-*: *hi-ta-n* 'in you (sg.)', *hi-ta-ra(t)* 'to you', *hi-ta-rik* 'from you', etc. In addition, a third procedure —actually also present in some HN. varieties and authors— is to use the postposition *bait(h)a-*, either on the abs. or on the gen.: *hi(re) bait(h)a-n*, *hi(re) bait(h)a-ra(t)*, *hi(re) bait(h)a-rik*, etc. This range of possibilities is variously employed depending on the noun declension case. For the ines. (see *OEH*: 'hi'), in old texts the most frequent procedure is *hire bait(h)an*. For the all., the most frequently employed procedure is *hiregana(t)* in Lç. and Old L. auhors. For the abl., in old texts the dominant form is *hireganik* (as in Lç. and E.Cib.). However, Dch. uses *zu-tan* 'in you (sg., formal)' but *zu-gana* 'to you', *zuie-ganik* 'from you (pl.)' (also *ni baita-n* 'in me', and *ni baita-ra* 'to me'). In this author, then, *-ta* has penetrated only into the ines., whereas the all. and abl. are formed by *-gan-a/-ik*. Unlike in Old L. authors, *-gan-* is attached to the abs., not to the gen. A similar pattern is found in Oih. and Tt.[36]

Out of all these procedures, the *hi-gan* type is the most widespread in old texts, and appears to be common. The *hi-re-gan* type, which is present both in

35 And the same occurs with the western sociative *-gaz*, which in Laz. is attached either to the bare form of the personal pronoun (*zu-gaz* 'with you (sg., formal)') or to the genitive form (*zure-gaz*). In the later tradition, however, the former possibility is almost exclusively used.

36 Both of these, however, have some cases like Oih. *ene-gan-a* 'to me' and Tt. *ene-gan-ik* 'from me'. It appears, then, that the 1st p. sg. has been the readiest to attach *-gan-* to the genitive.

Old L (actually as the only possible type with *-gan-*), and in Laz. (in free variation with the *hi-gan* type), is most probably a central innovation.

7.3.3. Weak vs. strong forms

The pronoun forms discussed in §7.3.2 are the weak or neutral ones. But all dialects have a productive way of suffixing the pronominal root by means of the deictic element *haur*. I shall call this the 1st personal pronoun suffixing process (henceforth 1st PPSP). By means of this process, the combination of *ni*, *hi*, *gu*, *zu*, *zu(ek)* + *haur* gives rise to *ni-haur*, etc., by and large in LN, S, Sal and R. In the rest of the dialects, these evolved phonetically, in the abs. case, into *neu*, *heu*, *geu*, *zeu*, *zeuek*, although, as will be shown below, these forms do not actually appear in some central areas or authors. The other cases are built by attaching the case marker to *-haur*. Thus, for the erg. we have *ni-haur-k* in Dch. (East.-LN.), although, from the 19th century onwards, *ni-haur-ek* is more frequent in this area (see *OEH*: 'nihaur'). To the west of this area, *ni-haur-k*, etc., evolves into *neu-k*, etc., but the step with the rhotic is attested in Old L (Ax. *neur-k*, *zeur-k*), Laz. (*neur-k*, *eur-k* 'you (sg.)'), and in a few other authors (see *CA*).

The dat. resulting from the 1st PPSP is not attested, in the phonetically non-reduced form, until Mst. *ni-haur-i*, *gi-haur-i*, *zi-haur-i* (at least, according to the *CA*; see also *EMH*: 34). Phonetically reduced forms are not very abundant. In the *CA*, the only two attestations in the 16th century are Laz. *neuri* and Lç. *euri* 'to you (sg.)', and, in the first half of the 17th century, E.Cib. *neuri*. As for the gen., in the *CA* the first attestation of a non-reduced form is Oih. *ni-haur-en*, *(h)i-haur-en*, and Bp. *gi-haur-en*. Note that the gen. marker here, unlike in the rest of the cases with pronouns, is the modern *-en*. Hence, these are probably late formations upon abs. *ni-haur*, etc. The reduced forms *neur-e*, *(h)eur-e*, *geur-e*, *zeur-e*, *zeu-en* (< **ni-haur-e*, etc.) are found profusely in the oldest texts of every dialect, and they are in complementary distribution with the corresponding weak *en-e*, *hi-r-e*, *gu-r-e*, *zu-r-e*, *zu-en*, according to the Linschmann-Aresti Law (see §7.3.4), together with the 3rd p. gen. pronouns.

There is another process, which I shall call the 2nd personal pronoun suffixing process (henceforth 2nd PPSP), the outcomes of which do not appear in all dialects. To be more exact, they appear neither in Biscayan nor, with a couple of exceptions, in those varieties where the outcomes of the 1st PPSP are not phonetically reduced. By this process, another demonstrative, in this case *hau(r)* or *hori*, is attached to the result of the 1st PPSP. These forms are made by inflecting the corresponding demonstrative in its usual way (see Table 7.1). Thus, Laz. has *neurau* 'I myself (abs.)', which is *-au* attached to *neur-*, < *ni-haur*. But he also

has the corresponding erg. form *neur-onek*, with the erg. form of the demonstrative, and the dat. form *neur-oni*. Similarly, he has *zeur-ori* 'you (sg. formal, abs.)' (*zu-haur* > *zeur-* + *-ori*), and the corresponding erg. *zeur-orrek* (*zu-haur* > *zeur-* + *-orrek*). If we compare these to Dch. *i-haur-orrek* 'you (sg., familiar, erg.)' and *zu-haur-orrek* 'you (sg., erg.)' (see *EMH*: 34), there seems to be an underlying procedure, which in Dch. still remains at an analytical stage, undisturbed by phonetic evolution. First comes the pronoun, then *haur*, interpreted as an adverbial element of proximity (hence its lack of inflection, as is the case in the 2nd PPSP) and unspecified for 1st or 2nd p., and finally, attached to the result of the 1st PPSP (in its bare form), *hau(r)* for 1st p. and *hori* for 2nd p. pronouns. These are interpreted as demonstratives, and as such are inflected. Disregarding 2nd p. pl. forms —for which satisfactory examples are hard to find—, this underlying procedure is shown in Table 7.6 by means of some representative attestations.

Table 7.6. Weak (or neutral) pronouns, 1st and 2nd Personal Pronoun Suffixing Processes

	Weak pron.	**1st PPSP** (+ adverbial *haur*)	**2nd PPSP** (+ 1st p. *haur*, *hon-ek* ... / 2nd p. *hori*, *horr-ek* ...)
Abs.	*ni*	*ni-haur* > *neu(r)*	**ni-haur-haur* > Ag. *nerau*, Ax. *neror*, Laz. *neurau,* Lç. *neuror*
	hi	*(h)i-haur* > *(h)eu(r)*	**(h)i-haur-hori* > E.Cib. *herori*
	gu	*gu-haur* > *geu(r)*	**gu-haur-hauek* > Añ. Lç. Pouv. *geurok*
	zu	*zu-haur* > *zeu(r)*	**zu-haur-hori* > Ag. Ax. *zerori*, Ber. Laz. *zeurori*
Erg.	*ni-k*	*ni-haur-k* > *neu(r)k*	**ni-haur-honek* > Laz. *neuronek*, Ax. E.Cib. *neronek*
	hi-k	*(h)i-haur-k* > *(h)eu(r)k*	**(h)i-haur-horrek* > Mend. *erorrek*, Lç. *eurorrek*, Dch. *ihaurorrek*
	gu-k	*gu-haur-k* > *geu(r)k*	**gu-haur-honek* > Haran. *geronek*, Pouv. *geurok*
	zu-k	*zu-haur-k* > *zeu(r)k*	**zu-haur-horrek* > Ax. *zerorrek*, Dch. *zuhaurorrek*
Dat.	*ni-r-i*	*ni-haur-i* > *ne(u)ri*	**ni-haur-honi* > Laz. L.Elc. *neuroni*, Ax. *(niri) neroni*
	hi-r-i	*(h)i-haur-i* > *(h)euri*	**(h)i-haur-horri* > Lç. *eurorri*
	gu-r-i	*gu-haur-i* > *geuri*	**gu-haur-honi* > Haran. (abs.!) *geroni*, Añ. *geuroi* (but here < dat. pl. *-oi*)
	zu-r-i	*zu-haur-i* > *zeuri*	**zu-haur-horri* > Ax. *zerorri*
Gen.	*en-e*	*ni-haur-e* > *neure*	**ni-haur-honen* > Laz. *neuronen*, L.Elc. Mend. *neronen*
	hi-r-e	*(h)i-haur-e* > *(h)eure*	**(h)i-haur-horren* > Azk. *herorren* (20th c.)
	gu-r-e	*gu-haur-e* > *geure*	**gu-haur-hauen* > Lç. Pouv. *geuron*, Duv. *geronen* (19th c.) (but here < gen. pl. *-on*)
	zu-r-e	*zu-haur-e* > *zeure*	**zu-haur-horren* > Pouv. *zerorren*

It must be highlighted that Table 7.6 is designed for didactic purposes. Some forms, and particularly some of those with an asterisk in the right-hand column, are theoretical constructions employed to fill all slots rather than conceivable reconstructions. Thus, a form like Pouv. *geur-on* does not come from **gu-haur-hauen*, but most probably from the *-on* of the noun morphology attached to an evolved form of **gu-haur-*. Similarly, although the abs. form of the 1st p. sg. pronoun of the 2nd PPSP points to **ni-haur-haur*, the historical attested forms corresponding to it must have been formed not from **ni-haur-haur*, but most probably from the attachment of *-aur* (> west. *-au*, Old L. *-or*) to a phonetically evolved form of the outcome of the 1st p. sg. abs. pronoun resulting from the 1st PPSP (i.e. *ni-haur-* > *neur-*, or even > *ner-*, + *-aur* > Laz. *neurau*, Lç. *neuror*, Ag. *nerau*, Ax. *neror*, etc.).

What I rather wish to show is that, as I see it, the 1st PPSP forms were already formed in CB, and that in fact the mechanism underlying the 2nd PPSP was also latent in it, although the fulfillment of this did not occur until a dialectal period. This mechanism consisted in the correspondence *-au(r)* → 1st p., *-ori* → 2nd p. It must also be pointed out, however, that, already in the historical period, this correspondence has broken down in a part of the continent, as, at least in Lç. and Old L. authors, some crossing appears to take place. Thus, Lç. has erg. *neur-orrek*, and E.Cib. has erg. *zer-onek* (also *zer-orrek* and *zu-k zer-orrek*). In these cases, the demonstrative is not attached to the person to which it corresponds, and this tendency increases in Labourdin in the following centuries (see *EMH*: 34-37).[37]

For the diachronist, however, the most interesting issue is how and why the 1st and 2nd PPSPs have taken place at all, as well as in what order the slots involved in them —along both the case axis and the 1st PPSP → 2nd PPSP axis— may have arisen. My prediction is the following: the 1st PPSP started off in the genitive. In fact, this slot (i.e., the gen. of the 1st PPSP) contains both formal and semantic features which make it different from the rest of the slots. On formal

37 Another interesting phenomenon is that, on some occasions, the dat. of the pronoun has become abs. Thus, *ner-oni* and *ger-oni* look morphologically like datives, but semantically they are absolutives, in most of the cases, in continental authors. Thus, we have, for example, *neroni* in E.Cib., Pouv., Gç., etc. E.Cib. also has *zu zer-oni*, where the same procedure has spread to the 2nd p. For *geroni* (abs.), the first attestation in the *CA* is Haran. (also Labourdin, but in the following century). According to the *OEH* (see 'neroni', 'geroni'), these forms —as absolutives— have also been used in Guipuscoan from the 19th century onwards. Unlike Azkue (*MV*: 441), I do not think that these forms are archaic. They might be datives reinterpreted as absolutives in positions where those datives were non-canonical subjects. Such processes are not unusual cross-linguistically, as Eng. *me too* shows.

grounds, gen. forms such as *neure* are the only ones which, in addition to being general to all dialects, appear in a phonetically reduced form in all of them (**ni-haur-e* > *neure*, etc.), which may be, by the way, a further indication of their antiquity. On semantic grounds, as will be shown in §7.3.4 the appearance of *ene* or *neure*, etc., in a given context, is syntactically conditioned, whereas in all the other cases the conditioning which triggers the alternation between the neutral, the strong and the hyperstrong forms (e.g. erg. *ni-k* / *neu(r)-k* / *ne(u)r-onek*) is a pragmatic one. It is clear that the former conditioning chronologically precedes the latter, as the historical evolution of gen. forms in the west shows.

The following step appears to be the extension of the 1st PPSP to other cases, now under a pragmatic conditioning. To which one(s) did it extend first? And in which case(s) did the 2nd PPSP begin? I believe that the research of this issue is still in its infancy. A chronology of the rise of the slots of Table 7.6 depends on a careful study, still to be made, of the dialectal distribution and historical development of the attestations corresponding to each of those slots. Such a discussion cannot be undertaken here, but I would like to focus, as an example of potential research, on the abs. slots of the 1st and 2nd PPSPs.

Examples (1a-c) from Axular show both the empty nature of the abs. slot corresponding to the 1st PPSP —i.e. the inexistence of forms like *neu(r)* in this author— and some pragmatic nuances underlying the alternations between neutral, strong and hyperstrong forms:

(1) a. *Eta baldin zu hala behar-tzen ba-zaituzte,* (Ax., 177)
and if you (sg.) so oblige-IPFV if-AUX

zeur-k du-zu falta, zerori zara kausa
you(strong)-ERG AUX-2SG fault you(hyperstrong).ABS are reason
'If they so oblige you, YOU are at fault (lit. you have the fault), YOU are the reason'

b. *Mundu herri guzti-ak (...) ezin egotz zaitzake ifernu-ra.*
world land whole-ERG.SG can.not throw AUX hell-ALL.SG

Baina (...) zerorrek egotz-ten du-zu
but you(hyperstrong).ERG throw-IPFV AUX-2SG

zeure buru-a (Ax., 399)
your(strong) head-ART
'The whole world cannot throw you to hell. But YOU throw yourself'

c. *Paga* *eza-zu* *zerorrek* *zeur-k* *zor*
pay AUX-2SG you(hyperstrong).ERG you(strong)-ERG owe

du-zu-n-az *geroztik* (Ax., 162)
AUX-2SG-RLV-INSTR.SG for, since
'Pay it yourself, for YOU owe it'

In (1a), we find two coordinated contexts in which the 2nd p. sg. pronoun must be focalized, first in the erg., then in the abs. Yet the forms used in those identical contexts are outcomes of different processes: the strong 2nd p. erg. sg. *zeur-k* derives from the 1st PPSP, whereas the hyperstrong 2nd p. abs. sg. pronoun *zer-ori* derives from the 2nd PPSP. (1b-c) show that the absence of a hyperstrong 2nd p. erg. sg. form in (1a) is not due to its inexistence in Ax. In fact, in (1c) both erg. *zeur-k* (< 1st PPSP) and erg. *zer-orrek* (< 2nd PPSP) appear side by side, their function being very similar. A further possibility present in Axular is the duplicated *zu-k zer-orrek*. This amounts to three focalizing variants of the erg. of the 2nd p. sg. pronoun, in addition to the neutral variant. All of them mark the focus of the sentence (called *galdegaia* in Basque grammar), and are often located immediately before the verb. But as Rebuschi (1995: 325) admits, it is extremely difficult to make out which semantic or pragmatic nuance, if there is one, could account for the preference of one over the other two options.

Leaving *zu-k zer-orrek* aside, in Ax. we find, then, a threefold opposition for the erg. (*zu-k* → *zeur-k* → *zer-orrek*), whereas for the abs. the opposition is a twofold one (*zu* → ***zeu(r)* → *zer-ori*).[38] It appears that, along the process neutral pronoun form → 1st PPSP → 2nd PPSP, the erg. has grammaticalized forms at the three stages, whereas, in the abs., once *zu-haur* > *ze(u)r* was created (1st PPSP), this did not grammaticalize but jumped directly to the 2nd PPSP.

Inversely, in Biscayan (and Guipuscoan until the 18th century; see *OEH*, although there are actually very few earlier texts in G), there is no *ne(u)rau*, *ze(u)rori*, or any other abs. form resulting from the 2nd PPSP. By contrast, abs. forms resulting from the 1st PPSP such as *neu*, *geu*, *zeu*, etc., appear from the earliest texts. Here, then, we have the twofold opposition abs. *zu* → *zeu(r)* → ***zer-ori*. As in other aspects, here again Lazarraga reveals himself as a bridge between the west and the center/east, as he attests forms of the three stages, at least for the 2nd p. sg. (formal): abs. *zu* → *zeur* → *zeur-ori*.

In fact, if we dispense with modern literary works, in Biscayan there are no forms resulting from the 2nd PPSP for any case. It is interesting to compare this to

38 The same is true for the rest of the persons. In the 1st p. sg., for instance, we have Ax. *ni* → ***neu(r)* → *neror* (~ Lç. *neuror*).

the situation in the east, as, leaving aside the two cases by Dechepare mentioned above (*i-haur-orrek* and *zu-haur-orrek*, both of them 2nd p. sg. (familiar and formal) erg.), I know of no further attestations of forms resulting from the 2nd PPSP in those dialects where the forms resulting from the 1st PPSP do not undergo phonetic reduction (cf. *CA*). By contrast, as Blanca Urgell points out to me, Old Souletin authors do show a vigorous and productive system of 1st PPSP forms for all the cases. These forms (abs. *ni-haur*, erg. *ni-haur-k*, etc.) have been sometimes explained as new pronouns rebuilt upon the neutral *ni*. This can be valid for the gen. forms *ni-haur-en*, etc., the secondary nature of which seems borne out by the fact that the genitive form is the modern one (*-en*, not *-e*). But, for the rest of the forms, I consider this assumption unnecessary.

Instead, I feel more inclined to think that, gen. *ni-haur-en*, etc., left aside, the rest of the far eastern *ni-haur* forms of the 1st PPSP were formed already in CB —understood in a wider sense than the purely chronological one—, just as the corresponding ones of other dialects. Next, in a dialectal period, these forms underwent phonetic reduction into *neu(r)*, *neu(r)-k*, etc., in the rest of the proto-dialects. This disfigurement brought about the need of a recharacterization of their strong nature, which was expressed by the 2nd PPSP (→ *neur-au(r)*, *neur-onek*, etc.). This process spread centrifugally to the west and the east, but did not reach Biscayan, and did not succeed in the far eastern dialects either, which got by with a twofold opposition (*ni / ni-haur*, etc.).[39]

Turning now to old western texts, in these the function of strong forms is the same as in that of (1a-c): to highlight the focal prominence of the pronoun, as in the following example by Barrutia:

(2)	*Zeu*	*zara*	*ene*	*Xaun-a*	(Barrutia, *Actto* 49)
	you (sg.)	are	my	Lord-ART	
	'YOU are my Lord'				

In the later western tradition, however, all strong forms have tended to become increasingly less marked, so that sometimes it becomes extremely difficult to predict which form, the weak or the strong, should fit into a given context. This has gone beyond the western borders. Thus, in (3), taken from a grammar

39 A final question is why gen. forms underwent phonetic reduction even in the dialects where the rest of the cases did not. I have mentioned their greater antiquity as a possible factor, but, if all of them were formed by CB, perhaps a more powerful factor —pointed out to me by Borja Ariztimuño— was the fact that genitive forms are not nominal heads, but dependents and hence clitic-like. In addition, the rest of the 1st PPSP forms (*ni-haur*, *ni-haur-k*, etc.) are emphatic, which entails focal prominence, and hence a more "resistant" pronunciation.

of SB, a strong or duplicated form like *ni-k neu-k* can give up its focus- or rheme-marking role, and move on to mark the theme (called *mintzagaia* in Basque grammar):

(3) *Ni-k* *neu-k,* *ez* *nioke* *ezer* *esan-go*
I-ERG I(strong)-ERG not AUX anything say-PROSP
'As far as I am concerned, I wouldn't tell him anything'
[Taken from Euskaltzaindia (2002: 57)]

In eastern dialects, however, even today *ni-k neu-k* could never have such a meaning. Thus, a similar but affirmative sentence like *Ni-k neu-k esango nioke* could only be interpreted as 'I MYSELF would tell him', or even 'I ALONE would tell him'.

As a last (in fact recent) step of the demarking process of strong forms developed in the west, some Biscayan varieties have generalized these (abs. *neu*, erg. *neu-k*, etc.) for any context, thus causing neutral forms to disappear. This issue, however, proves somewhat slippery in the descriptions of Biscayan varieties, as the inner tendency often clashes with the influence of SB, and perhaps also of the prestigious Guipuscoan dialect in previous periods. To give but one example of such a clash, in their description of the Biscayan of Lekeitio, where the 2nd p. familiar treatment (cf. §9.6.2) has been lost among young generations, Hualde & Elordieta, G. & Elordieta, A. (1994: 103-104) give both sets (*ni*, *neu*... / *ni-k*, *neu-k*...) as possible, but note that the 2nd p. sg. familiar pronoun, still remembered by some old speakers, is exclusively reported as *eu* (the strong variant of *(h)i*), *eu-k*, etc.

7.3.4. The Linschmann-Aresti Law

This law predicts the appearance of *ene* or *neure*, along with the other weak or strong gen. pronouns (including 3rd p. sg. and pl.), in a given context, and is effective, to different degrees, in the old texts of every dialect. It can be formulated in the following way: in a given context, between 1st sg. *ene / neure*, 2nd sg. (fam.) *hire / (h)eure*, 3rd sg. *haren / bere*, 1st pl. *gure / geure*, 2nd sg. *zure / zeure*, 2nd pl. *zuen / zeuen*, and 3rd pl. *haien / bere*, the strong variant will be chosen if the person corresponding to that gen. (e.g. *neure etxe-a* 'my house', → 1st p. sg.) is indexed —i.e. has a (non-allocutive) (overt or covert) morpheme corresponding to it— in the finite verb of the corresponding clause, whether main or embedded. In other words, if it is a core argument of that clause. Here are some examples:

(4) a. *Eriden du-t Dabid Iese-ren seme-a (...), [zein-ek*
find AUX-1SG p.n. p.n.-GEN son-ART REL.PRON-ERG

egin bai-titu ene nahi guzi-ak] (Lç., *Deeds* XIII, 22)
do CPL-AUX my desire all-ABS.PL
'I have found David, Jesse's son, who will fulfill my desires'

a'. *Zuek du-zue ezagutu-ren [ezen ni na-iz-ela neure*
you. PL AUX-2PL know-PROSP that I 1SG-AUX-CPL my

Aita baitha-n] (Lç., *John* XIV, 20)
father particle-INES.SG
'You (pl.) will know that I am in my Father'

b. *Non da zure anaia Abel?* (Ax., 337)
where is your brother Abel
'Where is your brother Abel?'

b'. *Gibela eza-zu zeure oin-a adiskide-aren etxe-tik*
remove AUX-2SG your foot-ART friend-GEN.SG house-ABL.SG
'Remove your foot from the friend's house' (Ax., 27)

c. *Ni aren min-ez il-te nin-za-n* (Laz., <1181v>)
I his pain-INSTR.INDEF die-IPFV 1SG-AUX-PST
'I died because of pain for her'

c'. *Igo ebe-n bere ugazaba-ren aposentu-ra*
go up AUX.3SG-PST his lord-GEN.SG room-ALL.SG
'(He) went upstairs, to his lord's room' (Laz., <1143v>)

The embedding can be more complex than that of (4a, a'), but the details as to how the law behaves in such cases, as well as what happens when the corresponding verb is non-finite, or is omitted, or what happens with vocatives, must be skipped here (see Sarasola 1980, Rebuschi 1995). I shall briefly take up two issues: how it worked in the 16th-17th centuries together with how it has evolved since then, and how it may have arisen.

In old texts, the law applies for all persons in Old Labourdin (it is systematic in Ax. and E.Cib.). In Dch., Lç. and Tt., the strong forms *geure*, *zeure*, *zeuen*

have been substituted by the corresponding weak ones.[40] In Old HN. texts and authors (Ber. and Eliç.), the law also applies to all persons except to the 2nd p. pl. In both Ber. and Eliç. there are, however, exceptions. In Ber., some fixed NPs such as *Iesu Kristo gure Iaun-a* 'Jesus Christ our Lord' always have weak gen. pronouns, regardless of the corresponding verb (see Sarasola 1980).[41] In Laz. (Alavese), the law is almost always applied. In the scanty Guipuscoan texts of the 16th and 17th centuries, the law applies, but in Och.A. the law has fallen apart, as the strong forms have been basically generalized for any context. In Mend. and Ub., the law continues to apply for 1st p. pl. *gure / geren* and 2nd p. sg. *zure / zere*, but in the 1st p. sg. the only form is *nere* for any context. As for Biscayan, the law usually applies in Cap., although in some cases it has ceased to apply, especially because strong forms have already begun to sneak into contexts where the weak form is expected according to the law (cf. Arejita 2008: XV). This tendency increases in Amill., where strong forms sometimes appear in contexts where they are not expected in order to mark the focus of the sentence, or in vocatives. In some other cases, it is difficult to find any reason for this encroachment (cf. Arejita 1988: 113-114). In Ol., we can say that the law does not apply any longer (cf. Sarasola 1980: 440). Both weak and strong forms can appear, but there is no way to predict which one will appear in a given context.

The details as to how the law falls apart have not been thoroughly researched, except for in continental authors (see Rebuschi 1995).[42] We can briefly summarize this process in the continent in the following way. 1st and 2nd p. gen. strong pronouns disappear in favor of the weak ones (especially from the 18th century onwards, see *OEH*: 'neure', etc.), but the 3rd p. gen. forms preserve their complementary distribution (at least until very recently; cf. Trask *GB*: 153): *bere*, *beren* if they are indexed (syntactically and semantically) on the corresponding verb, *haren*, *haien* otherwise.[43] A similar decline of 1st and 2nd p. gen. strong pronouns occurs in the HN. and G. traditions. In the former, the Lin-

40 In addition, Rebuschi (1995: 343) observes that, in Dch., the gen. pronouns within NPs which are subjects are always the weak ones.

41 Sarasola (1980) leaves the 3rd p. pronouns out of the law, but according to the *OEH* (see 'bere') the use of *bere* is strictly anaphoric —hence according to the rule— in all old texts outside Biscayan and in much of the later tradition in most dialects. As for High Navarrese, in Ber. *bere* is only anaphoric (hence, according to the law), whereas in Eliç. and L.Elc. there is hesitation as to its use.

42 For B, cf. Eguzkitza's (1995) short but substantial paper.

43 An intermediate step in the development of the law in the continent is attested in the text K&B (ca. middle of the 18th century; cf. Padilla-Moyano 2011: 83-85), where the law is still applied with the 1st p. sg. (*ene* / *neure*), in addition to the 3rd persons.

schmann-Aresti Law distribution also shows hesitation in the 3rd p. forms from Ber., where the tendency is for *bere* to encroach on non-anaphoric contexts, as is the case elsewhere in peninsular dialects (especially in Biscayan). In the latter, the Linschmann-Aresti Law disappears by generalizing one variant (the strong one in the 1st p. sg., *nere* < *neure*; the weak one in the other 1st and 2nd persons), but in the 3rd p. realm, *bere* is usually restricted to anaphoric contexts until the 20th century (see *OEH*: 'bere'). In this century, *bere* widens its scope. In Biscayan, when the Linschmann-Aresti Law falls apart, the 1st and 2nd p. gen. strong pronouns couple with the rest of the strong pronouns, and fall into the same distribution based on pragmatic factors. Sometimes they can focalize the corresponding NP, usually before the verb, as in the following example:

(5) *Zeure-ak* *dira* *gure* *biotz-ak* (Barr., *Actto* 396)
your(strong)-ABS.PL are our(weak) heart-ABS.PL
'YOURS are our hearts'

Note that in (5) the Linschmann-Aresti Law would predict weak gen. pronouns in both *zeure-ak* and *gure*. However, the law has ceased to apply, and now the conditioning is different: since the former is the rheme of the clause, it takes the strong form. This never works as a rule, though, and as I described earlier, in Biscayan, strong forms develop into a demarking process, so that it is really difficult to predict whether the weak form or the strong one fits better into a given context. The 3rd p. *bere* has tended to become a 3rd p. gen. pronoun, fencing *aren* into deictic contexts (see, however, Eguzkitza 1995: 428).

As for the genesis of the law, it should first be noted that the Linschmann-Aresti Law must undoubtedly be a feature of Common Basque, most probably effective for all persons (as in Axular). How it arose can be discussed exclusively on internal grounds. If we leave the instr. aside, the gen. is the only grammatical case of Basque which is not indexed on the finite verb, and which often appears inside phrases with a core (abs., erg. or dat.) syntactic function. Furthermore, it is the only grammatical case which can anaphorically —in the narrow, generative sense— refer to an antecedent of the same clause.[44] However, the

44 In other words, out of the abs., erg. and dat., none can refer back to any of the other three in the same clause. Such a phenomenon occurs in languages with reflexive pronouns like English (e.g. *He$_i$ saw himself$_i$ / him$_j$ in the mirror*). But in Basque, the reflexive pronoun is the periphrasis *neure, heure, bere buru-a*, etc., lit. 'my, your, his/her head'. For instance: *Neure$_i$ buru-a ikus-i du-t$_i$ ispilu-an* 'I have seen myself (lit. my head) in the mirror'. Hence, the reflexive or anaphoric element is contained not in the abs., but in the gen. Similarly, the reciprocal pronoun in Basque is *elkar* (probably < **(h)ark-(h)ar*, lit.

consequence of this is different for the 3rd p. and for the other persons. If there is no embedding, the Linschmann-Aresti Law provides meaningful, contrastive oppositions only when the corresponding gen. is 3rd p. (*Peru-k bere* / *haren liburu-a aurki-tu du* 'Peru has found his own / somebody else's book'). If it is 1st or 2nd p., the information given by the choice of *ene* or *neure*, etc., is not contrastive.[45] Furthermore, the phenomenon with the 3rd person has parallels in other languages, such as e.g. in Latin (*suus* vs. *eius*), but with 1st and 2nd persons it is more unusual. In view of this, perhaps the set of oppositions arose precisely in the 3rd p., and later spread to the 1st and 2nd persons by means of some analogy, with the purpose of extending the reflexive/non-reflexive distinction throughout the whole paradigm, even if this was not contrastive in the 1st and 2nd persons. Once all persons had acquired a reflexive/non-reflexive opposition for gen. forms, this morphological opposition spread to other cases, but now with a pragmatic function, thus giving rise to the 1st PPSP described in §7.3.3.

7.3.5. 3rd person pronouns. Pronominal, anaphoric, and other uses

In §7.3.4, I have described the syntactic behavior of *bere*. As for its morphology, *ber-e* is the root *ber-* plus the archaic gen. *-e*. In old texts, this is the gen. of both the 3rd p. sg. and pl. The subcases which are formed by attaching a postposition to the gen. are also identical for the 3rd p. sg. and pl. (e.g. Ax. *ber-e-tzat* 'for him/her/it, for them').[46] From the 17th century onwards (*OEH*: 'beren'), in G and

'(s)he him/her'), which triggers 3rd p. sg. (!) object agreement on the verb. Hence, it cannot refer to the pl. subject and is not anaphoric.

45 I say "when there is no embedding" because, as Sarasola (1980: 442) points out, in cases like the following, 1st and 2nd p. weak / strong oppositions are also contrastive: Lç. (*Math* VI, 4) *eure Aita [sekretu-an dakusa-n]-a-k, rendatu-ren drau-k agerri-an* 'your Father, who sees in secret, will reward you'. The relative clause is after the modified NP, and hence is an apposition of 'your Father'. *Eure Aita* is the subject of the main clause here, and that is why *eure* is found. If we had *hire Aita ...* instead, then this would have to be interpreted as the direct object of *dakusa-n-a-k* 'who sees' (with no 2nd sg. (fam.) p. morpheme), the meaning being 'who sees your Father in secret, will reward you'.

46 The local cases of animates with *ber-*, are formed upon the gen. *ber-e-* in Laz. (e.g. Laz. *ber-e-gan-*'in him'), but in the B. and G. traditions the tendency is to form them upon the abs. (e.g. Cap. D.Zeg. *ber-a-gan* 'in him'), although there is hesitation. In the soc., this hesitation is already present in Laz. (two cases of *ber-e-gaz* 'with him/her', one case of *bera-gaz* 'idem'), and *bera-gaz* is practically the only form in the B. tradition (5 cases of *beregaz* (two of them in Laz.) / 782 of *beragaz* in the *CA*). In continental authors of the 16th and 17th centuries, there is also hesitation between *ber-e-gan-* and *bera-gan-* (the proportion being 15 / 5 in Ax., 25 / 2 in Lç. (*NT*), 2 / 4 in Tt. (*OHB*)).

HN the 3rd p. pl. form *ber-en* is attested. This form takes the modern gen. suffix *-en*, perhaps by analogy with *zu-en* 'your (pl.)', where *-en* entails plurality. In continental authors, there are scattered cases of *beren* from Lç., but its generalization does not take place until the 19th-20th centuries.[47]

This root *ber* can be related to the prefix found in *berr-ogei* '40' ← *hogei* '20', or even in Oih. *berr-egin* 'elegant, well done, (lit.) redone' ← *egin* 'to do', conveying the idea of 'twice', or 'again the same'. Additionally, when it is anteposed to the element to which it refers, it expresses the concept of 'same' in Souletin, as in S. *ber gizun-a* 'the same man'. As a postponed determiner, it exists in all dialects with the same meaning: Mog. *Apostol ber-ak* 'the same apostle (erg.)', Dch. *oren ber-ian* 'at the same hour', etc. In addition to this concept of 'idem', *ber-* as a postponed determiner can also express 'ipse'. In this case, the usual form is with *ber-* as apposition, taking the same case as the NP which it refers to: Cap. *premia-k ber-ak* 'the need itself (erg.)', Ax. *aita-ri ber-ari* 'to the father himself'.

However, it is usual in old texts (*OEH*: 'bera', 1st meaning) to find 'ipse' expressed without a case marker in the modified NP, in other words *Jainko ber-ak eman-a* 'given by God Himself', instead of *Jainko-ak ber-ak eman-a*. The former type is found e.g. in Dch., and in an award-winning poem from Pamplona-Iruñea (1609). It can also be found, along with the *Jainko-ak ber-ak* type, in Lç., E.Cib., Cap., Gç. and Bp. From the 18th century onwards, the *Jainko ber-ak* type hardly ever appears for 'ipse' and becomes specialized for 'idem', whereas the *Jainko-ak ber-ak* type expresses 'ipse'. The latter type, then, is probably an innovation, which in any case has prevailed in all dialects.

From the middle of the 17th century, the reduplicated form *berber-a* is attested both as a pronoun or as a determiner, expressing 'ipse' (*OEH*: 'berbera', 2nd and 4th meanings). As a determiner, and without a case marker in the modified NP, it already appears in Pouv. (*Iainko ber beraren presenzian* 'in the presence of God Himself'). As a pronoun, it is more frequently employed by peninsular authors. From the middle of the 18th century (idem, 3rd meaning), *berbera* as a determiner is also attested as 'idem', usually with no case marker on the modified NP (Haran. *Izpiritu berberean* 'in the same spirit'). It is employed somewhat more frequently by peninsular authors.

The idea of 'same' entailed by the root *ber-* must be the lexical source of a root which, as shown for the gen. *ber-e* in §7.3.4, subsequently becomes an independent pronoun. In Linschmann-Aresti contexts, gen. *ber-e* must already

47 As Rebuschi (1995: 321) points out, in some varieties in which the opposition *ber-e / ber-en* has come about, the derived soc. *ber-e-kin* 'with him/her/them' and dst. *ber-e-tzat* 'for him/her/them' continue to neutralize the number opposition.

have been an anaphoric pronoun in Common Basque, as the syntactic conditioning of that law is a common phenomenon (by 'anaphoric' I mean in the narrow sense that it must have a co-referent within the same clause). But upon the root *ber-* a whole paradigm has arisen, which is semantically pronominal —at least in western dialects— but which inflects like a normal noun: abs. sg. *ber-a*, erg. sg. *ber-ak*, dat. sg. *ber-ari*, gen. sg. *ber-aren*, etc., abs. pl. *ber-ak*, etc. This appears in all dialects from the beginning of the tradition. What varies across dialects is its semantic scope, since it is in competition with the deictic root *har-* (i.e. abs. sg. *hura*, erg. sg. *har-k*, dat. sg. *har-i*, etc.), which also has pronominal uses. In non-western dialects, *ber-a*, *ber-ak*, etc., is still intensive or emphatic, and is assumed to be in the same relation to *hura*, *har-k*, etc., as erg. *neu-k* to *ni-k*, dat. *neur-i* to *ni-r-i*, etc. The intensive or strong variants would emphasize the focal prominence of the pronoun.

However, in the 3rd person, the weak / strong opposition can potentially have another function which does not exist for 1st and 2nd person oppositions. Such a function would be that of distinguishing between two possible antecedents, not, as *ber-e* does, in the same clause (for it would be impossible in the case of a core argument, see §7.3.4), but in another clause of the same sentence (including the whole 'main clause + subordinate clause(s)' complex), or even in a previous sentence, or any previous part of the discourse. In authors and varieties where *bera*, *berak*, etc., has not encroached on contexts which originally belonged to *hura*, *hark*, etc. (i.e. the purely pronominal ones), *bera* has a certain anaphoric implication (and now 'anaphoric' must be understood in a broader sense). In the following sentences by Ax., the dat. *ber-ari* of the subordinate (completive) clauses in (6a, b) refers back to the subject of the main clauses. By contrast, in (6c) we find *hura* in the completive clause:

(6) a. *Hala dio San Agustin-ek [gertha-tzen zeika-la aldia batez,*
so says Saint Augustine-ERG happen-IPFV AUX-CPL once

ber-ari ere] (Ax., 30)
he-DAT.SG also
'So says Saint Augustine that it also happened to him once'

b. *{ ... } Iainko-ak, (...), errai-ten du [ber-ari eman zioe-la]*
God-ERG.SG say-IPFV AUX he-DAT.SG give-PFV AUX-CPL
'{Then, Martin gave that cloak to the poor and not to God. But in spite of that,}
God, (...), says that he gave it to him (i.e. to God)' (Ax., 151)

c. *Eta halako-ak { ... } uste du [ezen bertze*
and one like that-ERG.SG think AUX CPL other
guzti-ak ere hura bezala dire-la] (Ax., 253)
all-ABS.PL also he like are-CPL
'And someone like that { ... }, thinks that everybody else is like him'

Both cases of *ber-ari* are anaphoric inside the sentence, but, in addition to that, the one in (6b) is contrastive, for it helps clarify to whom Martin gave the cloak. However, *hura*, *hark*, etc., and *bera*, *berak*, etc., are not in complementary distribution in the same way as *har-en* and *ber-e* are, according to the Linschmann-Aresti Law (when it functions, as in Ax.), within the clause. In the cases other than the gen., it is not possible to predict which variant, whether the weak or the strong, will appear. In fact, in (6c) *hura* also has an antecedent in the main clause, but we find the weak variant instead.[48] In authors where the Linschmann-Aresti Law functions, the strong gen. *bera-ren* —along with its subcases *bera-rekin*, *bera-rentzat*, etc.— is, like *haren*, non-anaphoric, but emphatic, i.e. it involves some kind of focal prominence.

Therefore, in authors or varieties which are archaic on this point, the unmarked variant is *hura*, etc., and for strong variants to appear we need some triggering anaphoric (or emphatic) factor, although often it is impossible to predict where —i.e. how far back— and under what conditions, this factor suffices to trigger the strong variant.[49] In Old HN. texts, Laz. and Old B. texts, the situa-

48 The same holds true for old western texts. Consider this example in Cap.: *Bienabenturado-ak dira manso direan-ak, [zerren **aek** izan-go dira lurr-een jaba-ak]. Bienabenturado-ak dira negar egi-ten dabeen-ak, [zerren **eurak** consoladu-ak izan-go dira]. Bienabenturado-ak dira miserikordioso-ak, [zerren **aek** alkanzadu-ko dabee miserikordia]* 'Blessed are those who are meek, [since they will be the masters of the world]. Blessed are those who cry, [since they will be consoled]. Blessed are the merciful, [since they will reach mercy]'. In apparently identical contexts, Cap. uses the weak pronoun in the first and third sentences, but the strong one in the second.

49 Thus, the antecedent can be in another clause of the same sentence, as in *He$_i$ told me that he$_{i/j}$ was happy*. Or in a previous coordinated sentence, as in *John$_i$ came, and he$_{i/j}$ didn't give me the book*. Or in sentences separated by a pause, as in *I had plenty of friends$_i$. (...) Some of them$_{i/j}$ have died*. Even in contexts like the last one and in authors who are as archaic in this respect as Ax. or Dch., we occasionally find *ber-a*, etc. as the anaphora which refers back to the antecedent before the pause. In Dch. (XIII, 46), the poet is speaking about his enemies and, as an independent sentence, he says: *Haiek zer merexi dute-n, zuhaurorrek ikhus-zu* 'What they deserve, watch it yourself'. Several lines further down (XIII, 51), in the same context, with the same topic in mind, and starting a sentence —actually a stanza—, he says: *Ber-ak batzu hil dirade, ni are nago bizirik* 'Some of

tion is quite similar to the one described above, but in the attested history of western dialects —particularly B and G— *bera* becomes progressively less marked, pushing *hura* (B. *a*) into 3rd grade deictic meanings and becoming a fully-fledged 3rd p. pronoun. Thus, in Añ.'s translation of Ax.'s *Guero*, the Biscayan author sometimes translates Ax. *hura* as *bera* (see Urgell 2001: CXXXV).

The root *ber-* can be reinforced by a 1st grade or 2nd grade demonstrative. The forms created in this way (*ber-au*, *ber-ori*) have often acquired an anaphoric implication which refers back to the most immediate antecedent. If the 1st grade *(h)au* is attached, the resulting form is abs. sg. *ber-au*, erg. sg. *ber-onek*, etc. This can appear as a pronoun or as a determiner. In both cases, it can appear as an apposition, and, if so, it always entails some kind of emphasis. If it appears as an appositive determiner, the modified NP also has the 1st grade demonstrative attached in the same case form: Ax. *arrazoin hunek ber-onek* 'THIS reason, THIS VERY reason (erg.)'. *Berau* is attested (*OEH*: 'berau') in peninsular authors since Laz. (it is somewhat more frequent among the B. ones), but in continental authors it appears only in Ax. (*beror*), Pouv., Mst. (*beraur*), and a couple of others. A similar structure and range of possibilities exist with *ber-* plus the 2nd grade demonstrative *ber-ori*.[50] This is attested (*OEH*: 'berori') in the peninsula since 1622 (in a letter from Azpeitia), but in continental authors it is attested solely in E.Cib. (only after another demonstrative) and Pouv. *Ber-* plus the 3rd grade pl. demonstrative, *ber-aiek*, appears from the 18th-19th centuries in Souletin texts (*OEH*: 'beraiek'), and also in peninsular authors from the 20th century. This has become a 3rd p. pl. pronoun, also used in the standard language.

Finally, in Biscayan, Alavese and, to a lesser extent, Guipuscoan, we have another 3rd p. pl. pronoun: abs. *eur-ak*, erg. *eur-ak*, dat. *eur-ai*, gen. *eur-en*, etc. (G. *erak*, etc.; *OEH*: 'eurak'), already attested in Laz. It is a strong pronoun, so it behaves like *ber-ak*, etc. This pronoun can also be used in apposition with the meaning 'ipsi', the NP being in the same case form (Barr. *neskatx-ak eur-ak* 'the girls themselves (erg.)'). Its origin is unknown. It has the same vocalism as the strong pronouns of the other persons (*neure*, *(h)eure*, etc.). If it came from **beur-ak*, etc. (in fact *beure* 'his/her, their' is attested in E.Sar., a relatively late Labourdin author; cf. Irigoyen 1987 [1978]: 120), built from *bera* by analogy

---

them have died, whereas I am alive'. There are few such cases in Dch. —and this is the only one in abs. pl.—, but it would seem that from the increase of cases like these, the conversion of *bera*, *berak*, etc., into a 3rd p. pronoun should have taken place. But this has never occurred in continental (in fact, non-western) dialects, which use demonstratives as 3rd p. pronouns (cf. *HB*: 96).

50 Morphologically, this is identical to the 2nd p. pronoun of hyperformal treatment *ber-ori* (see §7.3.1). The latter must be a semantic deviation of the demonstrative. However, the earliest attestations of *berori*, those of Laz., are of 2nd p. hyperformal treatment.

with *neure*, etc., the fall of an initial **b*- would not be unparalleled (cf. *FHV*: 253). However, such an analogical process for the vocalism *à la indoeuropéenne* is unprecedented in the Basque of any period. Another possibility, suggested by Irigoyen (ibid.), is that **beur*- (in correlation to *neur*-, *heur*-, etc.) is the old form of *ber*-, so that it evolved into comm. *ber*- on one hand and into *eur*- in the west on the other. I consider this to be highly improbable.

7.4. Conclusion

In the preceding sections, the basic structure of demonstratives and personal pronouns in Basque has been presented. Due to the huge number of dialectal and historical forms in each category, only the most important ones for diachronic research have been mentioned. The idea which emerges is that, in the historical period, the demonstrative and personal pronoun systems have not considerably changed. However, by means of both the study of those changes —as well as dialectal comparison— and internal reconstruction, contemporary research can establish an overall picture of what the common language (i.e. Common Basque) was like in these respects, and, obviously from a more speculative perspective, what it may have been like in a pre-common phase (i.e. in Proto-Basque or Pre-Proto-Basque).

8. Non-Finite Verbal Morphology

Manuel Padilla-Moyano
(University of the Basque Country, UPV/EHU
/ Université Michel de Montaigne – Bordeaux 3)

8.1. Introduction*

Basque speakers can use verbs in two different ways: synthetic conjugation, consisting of a single word (e.g. *dakart* 'I am bringing it'), and analytic conjugation, also called periphrastic (e.g. *mintzatzen naiz* 'I speak', *ikusiko ditugu* 'we will see them'). Virtually, all verbs can be analytically conjugated, but only a limited number of them accept synthetic inflection. In fact, this number has been decreasing throughout historical Basque: from 68 verbs in the 16th century (Mounole 2011) to nearly a dozen in current Standard Basque, or even fewer in some varieties. However, it should not be inferred from this recessive evolution that synthetic conjugation is of greater antiquity.

Since Lafon's essential work (1980 [1944]) and the more recent contributions of Trask, this field has benefited in recent years from research into verbal grammaticalization (Aldai 2003), Proto-Basque reconstruction (Lakarra 1995 and subsequent works, Martínez-Areta 2006), the history of the gerund (Urgell 2006) and the general history of the verbal system (Mounole 2011). Such studies have greatly improved our knowledge of non-finite verbal forms. In this chapter we shall focus our attention on the history of Basque analytic verbs, their different parts and how they are combined. Each section will take present-day data as a starting point to describe the historical evolution of the elements studied and also, where possible, their prehistory.

* This work has received the support of the Basque Government by means of the predoctoral grant BFI 2010/018, MINECO's "Monumenta Linguae Vasconum (III-IV)" [= FFI2008-04516, FFI2012-37696] and UPV/EHU's UFI11/14. I would like to thank Bernard Oyharçabal and Blanca Urgell for constantly encouraging me, not only in this work, which has benefited from their interesting suggestions. I also thank Joseba Lakarra, Mikel Martínez-Areta and Manuel Padilla-Cruz for their linguistic advice and comments on this paper. Needless to say, I alone am responsible for any errors contained in this work.

8.2. Verbal periphrases in Basque

A Basque periphrastic verb consists of a non-finite form (main verb, henceforth referred to as MV) and an auxiliary verb (henceforth, AV), regardless of the eventual insertion of further elements such as modal particles. MV expresses aspectual information, while AV carries information about argument structure, tense and mood. In Basque periphrases, MV always appears in one of these forms: radical or participle (perfective, imperfective or future).[1] The combination of any participle with a given AV always produces an indicative form, while the use of the radical is reserved for the non-indicative moods.

Standard Basque or *euskara batua* has four auxiliaries, according to which periphrastic constructions can be classified. These are *izan* 'be' (indicative, intransitive), **edun* 'have' (indicative, transitive), **edin* 'become' (non-indicative, intransitive) and **ezan* 'obtain (?)' (non-indicative, transitive). To put it differently, there is one pair for the indicative and the other for marked moods: subjunctive, imperative, potential and conditional. This symmetrical division is the modern result of both diachronic evolution and geographical variation:

(1) *Ibil-tzen da*
walk-IPFV AUX.PRS
'(S)he is walking' or '(s)he walks'

(2) *Ibil dadin*
walk(RAD) AUX(IRR).PRS
'So that (s)he walks'

(3) *Ikus-ten dugu*
see-IPFV AUX.PRS
'We are seeing it' or 'we see it'

(4) *Ikus-i ditut*
see-PFV AUX.PRS
'I have seen them'

1 In this work, I employ the terminology used in Trask (1995b) with reference to non-finite verbal forms. For morphological reasons, however, Basque grammatical tradition uses *aditzoin* 'verbal base' instead of *radical*; *partizipio* 'participle' for only the perfective and future, and prefers *aditzizen* 'verbal noun' to *imperfective participle* or even *gerund*. I shall return to this terminological issue later.

(5) *Ikus* *genezan*
see(RAD) AUX(IRR).PST
'So that we might see it'

In addition, Biscayan and some varieties of Guipuscoan have *egin* 'do' —used as MV in all dialects and periods— instead of **ezan*,[2] and eastern varieties can show **iron* —perhaps the causative of an unknown verb (Gómez & Sainz 1995: 240)— but only for transitive verbs in the potential mood:[3]

(6) *Amai-tu* *da(g)izun*
finish-PFV AUX(IRR).PRS
'So that you may finish it'

(7) *Ikus* *giniro*
see(RAD) AUX(POT).HYP
'We would be able to see it'

The Basque verb can take three arguments: absolutive, dative and ergative. For tri-personal constructions of the indicative —that is, when the dative appears with transitive verbs— there is a set of forms which show significant dialectal variation. Eastern varieties exhibit forms like *derauko* '(s)he has it to him/her', of **eradun*, a causative of **edun*, whilst central and Standard Basque use forms like *dio*, where it is not certain if there is the presence of an auxiliary different from **edun*. In these forms, western Basque uses the dative flag *-ts-*: *deutso* (same gloss).

We have already seen how Basque periphrases are constructed with one of these forms: participle or radical. In the following section, I shall examine the morphology of non-finite verbal forms, as well as their diachronic evolution.

---

2 A complementary distribution of **ezan* (bi-personal conjugation) and *egin* (tri-personal conjugation) auxiliaries is also possible. Such a distribution was indeed present in Old Biscayan, Guipuscoan and Alavese, until the so-called literary dialects made their choices in favor of *egin* (Biscayan) or **ezan* (Guipuscoan) as transitive auxiliaries for marked moods.

3 **Iron* shows a recessive evolution in historical Basque. Thus, in early eastern texts (especially in Souletin, Roncalese and Salazarese) **iron* can occur as the only auxiliary verb for potential values in a whole range of possibilities, but at a given moment it is found in co-occurrence with **ezan*. Hence, in present-day Souletin **iron* remains only as a very defective verb in some forms like *diro* '(s)he can it' or *liro* '(s)he would be able to (do it)', **ezan* being the normal AV for all marked moods, as occurs in Standard Basque (Padilla-Moyano 2012).

8.3. Tense and Aspect suffixes of non-finite verb forms: origins and variation

Note that what we call the *imperfective participle* is formed from a gerund with the inessive marker *-n*; this gerund is likewise formed by adding a suffix to the verbal radical: RAD *ikus* 'see' → GER *ikuste* → IPFV PTCP *ikusten* 'seeing'. So, when dealing with non-finite verb forms in Basque, we must bear in mind the following tripartite plan: RADICAL – PARTICIPLE – GERUND. Such a limited set may at first sight appear to be too simple, but from a diachronic perspective it holds a remarkable richness, as I attempt to expound below.

8.3.1. About the verbal radical (-Ø) and participle[4]

8.3.1.1. A relative chronology of non-finite verbal forms

The verbal radical is undoubtedly the oldest form among non-finite verbs in Basque. While both the participle and gerund exhibit (both morphologically and functionally) clearly defined markers, the only affix which the radical is thought to take, **e-*, has an unknown function, even if some proposals treat it as a noun-deriving prefix (see Trask 1995b: 213), following the suggestion of Michelena (*FHV*: 423), for whom radicals were merely verbal nouns. Whatever its function might have been, the prefix **e-* is present in the most ancient verbs in Basque, although it has undergone different phonological evolutions: *eduki* 'have', *ekusi* (modern *ikusi* 'see', well recorded in Archaic Basque), Souletin *ebili* (common *ibili* 'walk'), *ekarri* 'bring', *jakin* 'know' (< **e-da-khin*). The mere presence of a prefix would be a sign of an ancient stage of the language, since typologically Basque is strongly suffixing, as SOV languages usually are.[5]

Without exhibiting remarkable dialectal variation in terms of its form or uses, the radical would have come, at least from the times before dialectal split, from what is called Common Basque (CB), as opposed to the modern unified Basque or *batua*. But it must be older than the participle suffix *-i*, and Bascologists widely accept that, from a morphological perspective, the verbal radical

4 In the absence of any specification, here *participle* will always mean 'perfective participle'. Otherwise, I shall use *imperfective participle* or *prospective participle.*

5 Indeed, such an **e-* was not the only prefix in Proto-Basque. Considering typological issues, the SOV word order is supposed to be one of the structural characteristics of Basque and, as Lakarra remarks (2006a: 564-5), this widely accepted idea has been too often uncritically assumed in PB comparison. Only Trask (1977), Gómez (1994), Gómez & Sainz (1995) and Lakarra have proposed a model for PB which differs from the canonical structure of Basque (see §10.4).

—and also *-i* participles— date from Pre-Proto-Basque. According to the CVC monosyllabic root theory (Lakarra 1995), which has enabled the recent advances in PB reconstruction, the most simple and ancient [*e-CVC] structure is still clear in many verbs: *e-kar* 'bring', *e-tor* 'come', *e-bil*, *e-kus*… As Martínez-Areta has realized (2006), there is no trace of [*e-hVC] structures in Basque verbs. Lakarra (2006a: 608) explains this absence by resorting to a relative chronology: for the beginning of the well-known evolution **th-*, **kh- > h-* —i.e. for the earliest centuries of our era— the [*e-CVC] rule would no longer have been productive;[6] otherwise Basque would lack [*e-T^hVC] structures like *ekhar* 'bring' or *ethor* 'come'. Consequently, ever since the change **T^h- > h-* occurred, the language has not created any new verb using the prefix **e*.[7]

On the other hand, both participle and gerund are clearly derived from the radical: *ikus → ikusi*, *ikuste*; *ibil → ibili*, *ibiltze*; *eman → eman*, *emate*. Note that the Basque perfective participle is the enunciation form for Basque verbs and, as mentioned above, modern speakers can easily deduce the radical as well as the derivation stem by simply removing the participle-forming suffix (except for the participles lacking any suffix, which are identical in form to the radical).

Finally, the gerund is the newest among the non-finite verbal forms in Basque, whereas both the radical and the participle date from PB. Its heterogeneous polymorphism, as well as its neatly described evolution, prompt us to place the origins of the gerund in a time after CB. More specifically —and following Urgell (2006: 927)—, it could be a medieval innovation, closely related to the development of the modern periphrastic conjugation and the consolidation of new ways of expressing verbal aspect.

8.3.1.2. Classification of Basque verbs

The participle has an interesting variety of endings, which yield the following classification of verbs, as proposed by Trask (1995b: 208).

6 When Lakarra considers this phonological change an *ante quem* element in order to date the verbal noun, he obviously deals with the PB forms of [*e-CVC] structure and derivates, which in the subsequent evolution of Basque have remained fossilized as verbal radicals.

7 Actually, a few new verbs with the prefix **e-* have appeared as a result of subsequent evolutions. Therefore, they do not involve any new root.

8.3.1.2.1. The -i *class*

There are several hundred verbs which take the suffix *-i* in the perfective participle. Virtually all of these verbs use the prefix **e-*. Therefore, *-i* is the most ancient among the participle markers, dating from PB. The radical of such a verb is precisely the form without the suffix: *erori* → *eror*; *ikusi* → *ikus*; *igorri* → *igor* 'send'; etc. Note that, at the eastern limit of Basque geography, Souletin has created what can be termed a "new *-i* class of verbs" with members such as *ezkhapi*, *konserbi* or *akhabi* from Gasc. *escapà* 'escape', *conserbà* 'preserve' and *acabà* 'finish', in co-occurrence with *ezkhapatü*, *konserbatü*, *akhabatü*. Obviously, these forms are entirely unrelated to the ancient participle suffix *-i*,[8] and their new *-i* cannot be removed, so that both the radical and the participle remain identical in form.

8.3.1.2.2. The -n *class*

Nearly one hundred verbs have the perfective participle with an apparent suffix *-n*: *izan* 'be', *egon* 'remain', *eman* 'give'. This *-n* is not removed when forming the radical. Again, all these verbs exhibit the prefix **e-*, which proves their antiquity. Trask (1990) argues that the verbs in this class were originally members of the *-i* class, the roots of which happened to end in *-n*. The regular loss of intervocalic *-n-* (*FHV*: 299-310) would have produced forms like *izai*, *egoi*, *emai* (attested in historical Basque) with the radicals *izan*, *egon*, *eman*. A further reorganization of these exceptional forms would be the origin of the modern *-n* class. This is a very suggestive hypothesis but, as Oyharçabal remarks (1998: 326), it leaves some loose ends, the most evident among them being the absence of such an *-i-* in the derived forms of some verbs of this class: *jan* 'eat' → ***jaite*, *edan* 'drink' → ***edaite*. However, Lakarra (2008a) proposes the evolution **-nin-* → *-i*.

8 This new *-i* would have its origin in Gascon verbal inflection, more specifically in the 1st person singular of the indicative present: *canti* 'I sing', *coneishi* 'I know', *sorteishi* 'I go out' from infinitives such as *cantar*, *conéisher*, *sortir*. Gascon largely influenced northern dialects of Basque until the presence of French became more evident, especially after the Revolution. For the whole Occitan area, this substitution process met its strongest resistance in Bearn (Brun 1923), and it is precisely the vigor of the Bearnese varieties which explains, by means of their vicinity, the fact that the influence of Gascon is especially relevant in Souletin.

8.3.1.2.3. The -tu *class*

This is the only productive class in historical Basque: every new verb entering the language falls within this class, with the only exceptions being a very small number of neologisms. It is commonly accepted that the *-tu* suffix is of Latin origin —although Trask (1995b: 212) suggested a PB. origin—, as shown by early loans such as *begiratu* 'look at, look after' < Lat. *vigilātu(m)*. Later, Basque borrowed Romance verbs in the same way, until *-tu* had replaced the ancient *-i* suffix. With the exception of a few verbs which seem to have been transferred into this class from other classes, no verb in this group takes the prefix **e-*. The suffix is *-tu*, which is absent from both the derivation stem and radical form.

8.3.1.2.4. The rest of the verbs

Finally, there are many verbs which cannot be included in any of the previous groups. For them, Trask created the "zero class": all of those verbs for which the perfective participle contains no removable suffix at all, so that the participle is identical in form to the stem and to the radical. This is a very heterogeneous group, with several identifiable subclasses of distinct origin. For these subclasses I shall follow Urgell (2006):

a) Verbs ending in *-ki* such as *ebaki* 'cut', *ireki* 'open' or *jaiki* 'stand up' or *ausiki* 'bite'. These verbs would have had an ancient suffix **-gi*, which very often appears —but not exclusively— in Biscayan forms: *ebagi*, *edegi* 'open', *jagi* 'stand up', *ausigi* 'bite'. According to Trask (1995b: 227), who agrees with Azkue, Schuchardt and Lafon in this respect, the morph *-ki* would be related to the dative flag *-ki-* that finite forms take in central dialects (e.g. *daramaki**o*** '(s)he brings it to him/her'). Lakarra (2006b) suggests that the morph *-ki* present in these verbs is the historical adverbializer *-ki*, but de Rijk (1998 [1995b]: 400-403) examined its origin without noticing such a connection. Most of these verbs take the prefix **e*, and none of them have any participle suffix, so that both the radical and perfective participle are identical in form.
b) The *-o* verbs, sometimes in alternation with *-an* forms: *igo* 'go up', *igaro* 'pass', *jaso* 'take' (also *igan*, *igaran*, *jasan*). All verbs in this subgroup have the prefix **e-*, and some of them present anomalies comparable to those exhibited by the *-n* class verbs. In fact, Trask proposes that these are "probably *-n* class verbs which have undergone some additional phonological development, now unrecoverable" (ibid.: 210).

c) Verbs ending in a vowel, without any participle suffix, and identical in form to their related adjectives: *busti* 'moisten', *bizi* 'live', *bete* 'full'. None of these verbs take the prefix **e-*, except *erre* 'burn'.
d) Verbs ending in a consonant, without any participle suffix. There is only one verb in this subclass: *hil* 'die'.
e) Verbs deriving from the allative case without any participle suffix. Thus, these verbs are identical in form to the allative case forms of the nouns: *atera* 'go out' (etymologically 'to the door'). No verb in this group exhibits the prefix **e-*. Note that any allative can be converted to a verb of the *-tu* class, but that only a handful of allatives form verbs of this subclass.
f) Some ancient loans which lack any participle suffix: *gorde*, *bota* and a few others. These verbs do not take the prefix **e-*. It is not always clear whether a verb should be best regarded as a member of this subclass or the adjective subclass.

Table 8.1. Classes of verbs

Verb ending	Examples	**e-* prefix	Ptcp. suffix
-i	*eror-i*, *ikus-i*	all	*-i*
-n	*egon*, *eman*	all	**-i* ?
-tu	*pentsa-tu*, *maita-tu*	none	*-tu*
-ki	*idoki*, *jaiki*	most of them	-
-o	*igo*, *jaso*	all	-
-V adjectives	*erre*, *busti*	none	-
-C adjectives	*hil*	none	-
allatives	*atera*	none	-
loans without suffix	*gorde*, *bota*	none	-

Concerning the participle suffixes, it can be argued that *-i* is the marker that the most ancient verbs took, regardless of whether the present-day *-n* class verbs were originally among them or not. As has been stated, these ancient verbs show the prefix **e-* and many of them also exhibit the canonical PB. root structure CVC: *e-kar*, *e-kus*, *e-gor*, *e-bil*… Lakarra (2006a) has explained a plethora of differing root structures in ancient verbs with the CVC theory by means of two reconstructed prefixes (**da-*, **ra-*) and certain phonological processes. These phonological processes may appear strange at first sight, but they have parallels in Bantu languages. Table 8.2 shows samples of these evolutions (since our aim is simply to illustrate phonological and word derivation processes, the meanings of the verbs are omitted).

Table 8.2. Etymology of several verbs dating back to PB

e-Ø- (non-marked)		*da-* (imperfective)		*(-)ra-* (causative-dative)	
**edan*	*> edan*	**e-da-dan*	*> j(a)an*	**e-ra-dan*	*> edan*
**eden*	*> eden*				
**eden(ki)*	*> edeki*				
**elen-i*	*> erein*	**e-da-len-i*	*> jarein*		
**edin(-o)*	*> jin, jo*	**e-da-din(-o)*	*> jaio*		
		**e-da-din(-ki)*	*>jaiki*	**e-ra-din-ki*	*> eraiki*
**elin(-o)*	*> (h)erio*	**e-da-lin(-o)*	*> jario*		
**edits*		**e-da-dits*	*> jaitsi*	**e-ra-dits*	*> eraitsi*
**elits*	*> eritsi*	**e-da-lits*	*> jaritsi*		
**edul*		**e-da-dul-ki*	*> jaulki*	**e-ra-dul*	*> irauli*
**edun*	*> *edun*	**e-da-dun*	*> jaun*	**e-ra-dun*	*> iraun*
			> ja(u)ntzi	**e-ra-duntzi*	*> era(u)ntzi*
**eduts*	*> eutsi*	**e-da-duts*	*> jautsi*	**e-ra-duts*	*> erausi*
**edutz*	*> eutzi*	**e-da-dutz*	*> jauzi*	**e-ra-dutz*	*> erauzi*
**egon*	*> egon*	**e-da-gon*	*> jagon*	**e-ra-gon*	*> eragon*
**egon(ki)*	*> egoki*				
**egotz*	*> egotzi*			**e-ra-gotz*	*> eragotzi*
**ekhin*	*> ekin*	**e-da-kin*	*> jakin*	**e-ra-kin*	*> irakin*
**enon*	*> eho(?)*	**e-da-non*	*> joan*	**e-ra-oan*	*> eroan eraman*
**esan*	*> esan*	**e-da-san*	*> jasan jaso*	**e-ra-san*	*> erasan eraso*
**eser*	*> eseri*	**e-da-sar*	*> jasarri*		
**ezar*	*> ezarri*	**e-da-zar*	*> jazarri*		
[From Lakarra (2006a: 584-585)]					

Apart from its use in participles, the suffix *-i* could also have given rise to adjectives such as *gazi* 'salted', *zuri* 'white', *gorri* 'red' or *hori* 'yellow', all of which are surviving forms in present-day Basque related to nouns or stems: *gatz* 'salt', *zur* 'wood', *gor-din*, *(h)or* 'dog'. Considering that radicals were PB. verbal nouns, a plausible hypothesis is that the suffix *-i* was productive for deriving participles from nouns, and that it consequently formed participles from radicals because radicals were nouns (Trask 1995b: 218).

Therefore, this suffix *-i* would have been productive over a long period of time, until the new participle marker *-tu*, entering from Latin-Romance, progressively excluded it. In addition to the Latin (such as *begiratu* 'look' < *vigilātum*) and Romance (such as *pentsatu* 'think' < *pensatu*) loans with which it entered, from a given moment long before the historical period of the language, the *-tu* suffix began to be combined with Basque roots too: *ajutu* (11th century), *aiçurtu* (1099). In historical Basque, *-tu* is already the only productive participle suffix:

in 1545, the suffix *-i* appears to be completely fossilized, and *-tu* verbs do not show the ancient prefix **e-*, which can easily be interpreted if we take the function of **e-* to be pre-Latin. The strength of the new participle suffix is evident not only from its productivity, but also from the fact that many *-i* verbs have joined the *-tu* class.

Nonetheless, in some western varieties, Spanish participles with *-ado* are regularly borrowed with the morph *-au* (from *-adu*): hence Spanish *bajado*, the participle of *bajar* 'go down', is borrowed as *bajau*. Spanish verbs in *-ido* are borrowed with *-idu*: hence Spanish *corrompido*, the participle of *corromper* 'corrupt' is borrowed as *corrompidu*. Landucci's dictionary and the recently discovered Lazarraga manuscript, both written in Alavese dialect, systematically attest this 16th century model for borrowing. This is a relatively recent development in western Basque, since earlier loans into Biscayan show the expected *-tu* —as in the other dialects— which offers, in fact, a valuable criterion for establishing a relative chronology of such Latin and Romance loans.

To put it simply, the extension of *-tu* is detrimental to *-i*. Different groups of examples related to this replacement process have been described in Urgell (2008: 7):

a) In addition to the loans with which *-tu* entered, certain CVC canonical Basque structure roots also began to take the new suffix: *bildu* 'collect', *galdu* 'lose', *hartu* 'take', *heldu* 'arrive', *kendu* 'remove', *lortu* 'obtain', *piztu* 'light', *saldu* 'sell', *sartu* 'enter', *sortu* 'born'… Several of them can be related to *-i* forms: *saldu* ~ *sari* (< **sal-i*), *piztu* ~ *bizi* (< **biz-i*), *galdu* ~ *gari* (< **gal-i*), *hartu* ~ *har-i-tu* (1545).
b) *-i* forms remaining as a noun or adjective: *neurri* 'measure' (used as a verb in 16th century texts) or *bihurri* 'windy, curving'.
c) The *-i* form being lost, but with some historical attestations: *aberatsi* 'become rich' (modern *aberastu*).
d) Pleonastic forms with an ancient *-i* base: *haritu*, *hotzitu*, *beltzitu* (SB. *hartu* 'take', *hoztu* 'cool' and *belztu* 'blacken').
e) Alternating pairs, depending on the variety and historical period, such as *ahantzi* / *ahaztu* 'forget', *irakurri* / *irakurtu* 'read', *bereizi* / *bereiztu* 'distinguish'.
f) *-n* class verbs which started to take *-tu* in central varieties, either replacing or not replacing the *-n*: *izandu* / *izatu* 'be', *egondu* / *egotu* 'stay', *iraundu* 'endure'.

8.3.1.3. Expansion of the participle in western Basque

As I have previously mentioned, in Standard Basque the verbal radical is used in periphrases with non-indicative moods, i.e. with the auxiliary **edin* for intransitive verbs and **ezan* for transitive ones. But in spoken language the radical, formerly common, remains only in the eastern dialects (Labourdin, both High and Low Navarrese, and Souletin), so that when an eastern Basque speaker says *erori da* '(s)he has fallen' but *eror daiteke* '(s)he can fall down', a western speaker makes no such distinction in MV: *erori da* and *eror**i** daiteke* (same gloss). For a detailed list of the uses of the verbal radical, see Trask (1995b: 214).

The loss of opposition between the radical and the participle in western Basque has taken place in the last centuries, since some fossilized idioms still attest a previous stage where the radical was in use, such as *or konpon!* ('sort it out as best you can!', where *(h)or* is 'there', and *konpon* the radical of *konpondu* 'arrange'). To a large extent, this neutralization process can be perfectly described by examining historical texts, following Lakarra (1996a: 185-8):

1) Verbs with participles which take the *-i* suffix would have lost their radical form in pre-historical Basque. This earliest loss occurred not only in western varieties, but would also have begun in northern Basque for one-syllable stems like *utz(i)* 'leave', even though the verbal radical has more recently undergone a reinforcement process in these varieties.
2) In Old Biscayan songs collected in the 16th century (but dating back to the 15th century), the opposition between radical and participle always occurs: *ezkon bekio* ('that (s)he marry him/her [jussive]', where *ezkon* is the radical of *ezkondu* 'marry'), *sar gaitezan* ('let's get in', where *sar* is the radical of *sartu* 'get in').
3) Participles did not appear in non-indicative moods until the 16th century, which could be an Alavese innovation.
4) From the 17th century onwards, the participle becomes the main form in non-indicative moods and starts to replace the radical, which will remain as an archaism. This process spread from Biscay to Guipuscoa and some areas of High Navarre.
5) The opposition has survived in word derivation, since the verbal radical is the normal derivation stem to which any suffix is regularly added: so, from *konpon* (participle *konpondu* 'solve') *konponketa* 'repair' (noun), *konponbide* 'solution, arrangement'. One-syllable verbs forming their participle with *-i* are the exception to this rule:[9] *hasiera* 'beginning' ← *has(i)*

9 In Biscayan and Guipuscoan, however, it is possible to derive gerunds from participles in *-tu* instead of deriving them from radicals, as in *agertute* ← *agertu* 'appear', and in some

'begin', *heziketa* 'education' ← *hez(i)* 'educate' or *utzikeria* 'negligence' ← *utz(i)* 'leave, abandon'.

Thus, if the participle is the current citation form of the Basque verb, it must be concluded that, at a previous stage, the verbal radical was the most evident reference to the verb, on the basis of three arguments: i) going back to the 16th century the verbal radical was a feature for all Basque dialects; ii) the radical is the form from which other non-finite verbal forms are derived; iii) ancient participles borrowed from Latin or Romance also created their own radical, unveiling an evident need.

8.3.1.4. Reinforcement of the radical in eastern Basque

If the radical has disappeared from western Basque, there is some evidence of an opposite tendency in eastern varieties. In effect, in eastern Basque, we can see a kind of reinforcement of the radical as a verbal form apart from the participle, and the more eastwards we go, the more the radical becomes marked.

It has been pointed out that in the *-tu* class verbs, the participle is obtained by adding the suffix *-tu* to the radical, and *vice versa*: in verbs borrowed as participles, the radical is formed by removing *-tu*. Nevertheless, eastern varieties usually form unmistakable radicals with *-t*: *lagunt* 'go with' (ptcp. *lagundu*) *zuzent* 'to direct' (ptcp. *zuzendu*), *deit* 'to call' (ptcp. *deitu*). Such verbal radicals, well-attested from 16th century texts, happen to be different in form not only with respect to the participle, but in many cases they also avoid homonymy with words of other categories (see Table 8.3).

Michelena (*FHV*: 423) noticed the addition of *-t* after certain radicals and, due to the de-nominal nature of some verbs involved in this phenomenon, suggested that the origin of such a *-t* could be related to well-known medieval phonological rules of word composition. This hypothesis implies great antiquity for radicals such as *hant* 'inflate' ← *handi* 'big' (ptcp. *hantu*), *ixilt* 'be quiet' ← *ixil* 'quiet' (ptcp. *isildu*) or *zilhet* 'permit' ← *zilegi* 'lawful' (ptcp. *zilhetu*). However it originated, *-t* would have become useful for distinguishing radicals from their correspondent nominal forms. Hence, we could propose an analogical extension of the *-t* ending forms to other verbs where the ancient word-formation rules would not explain radicals such as *xahat* 'purify' (ptcp. *xahatu*) or *deinat* ← *deinatu* 'condescend', and also to non-denominal verbs such as *irakurt* ← *irakurtu* 'read' and *sort* ← *sortu* 'create'.

---

varieties even derivation suffixes can do this: *apurtutzaile* 'that who breaks' ← *apurtu* 'break'. Every instance of derivation from the participle instead of from the radical is obviously an innovation, due to the neutralization process described above.

A confluent —but radically different!— phenomenon occurs with verbs in the *-ki* subgroup of the zero class, since eastern varieties show a tendency to construct radicals different from the participle by deleting the *-i* of *-ki*: *ideki* → *idek*, *ebaki* → *ebak*, *jaiki* → *jaik*. This procedure is a clear reinterpretation of such verbal forms being identified with the perfective participle suffix *-i*, which in eastern varieties —and also in Standard Basque— is regularly removed to obtain the radical: *ikusi*, *ikus* → *ideki*, *idek*. Lafitte (1979 [1944]: 203) gives a list of such forms: *aurdik*, *ausik*, *ebak*, *edek*, *edok*, *etxek*, *iradok*, *igurik*, *idek*, *jarraik*, and *atxik*, with the following seven verbs being exceptions to this rule: *eraiki*, *hunki*, *ilki*, *iraki*, *iraungi*, *jaiki* and *aurizki*.[10] But Lafitte misinterprets it when he says that forms such as *ebak* are original, since it is evident that verbs in this class are etymologically formed as [STEM + *-ki*], and not [STEM + *-k* + *-i*], so that the radicals of verbs in this group lacking *-i* must necessarily be an innovation; in fact, many of these forms appear around the 18th century, and the frequency of the phenomenon has been increasing throughout the last three centuries.

In a similar way, I should mention the reinterpretation of the causative *erazi* as having the suffix *-i*. In effect, a new radical *eraz* spreads from the 18th century, replacing the ancient radical *erazi*. Note that the old radical form also withdrew in peninsular dialects (necessarily before the process described in §8.3.1.3). This general loss of the ancient forms of the radical *arazi / erazi* explains why the modern standard language has chosen *araz*.

Finally, a few vowel-ending verbs with identical forms in the standard for the radical and participle exhibit different radicals in Souletin and Low Navarrese. Thus, for *busti* 'moisten', *bete* 'fulfill' and *gorde* 'hide' there exist the standard radicals *busti*, *bete* and *gorde* as well as the eastern radicals *busta*, *betha* and *gorda*. Such forms could also be linked to the need for differentiated radicals, as historically experienced by eastern Basque speakers.[11]

10 To my knowledge, *jaik* and its ancient causative *eraik* are both well attested in Souletin from the 18th century.

11 The same is applicable to *erre* 'burn', but the fact that the radical *erra* is also attested in an archaic Biscayan sentence (given in (23)) makes this issue more complicated. Lakarra has remarked to me that some old Biscayan texts also attest the participle *betatu*.

Table 8.3. Eastern radicals

	Participle	Common radical	Homonymous	Eastern radical
-tu class	*zuzendu* 'direct'	*zuzen*	*zuzen* (adj.)	*zuzent*, *xuxent*
	lagundu 'go with'	*lagun*	*lagun* (noun)	*lagunt*
	berdindu 'make equal'	*berdin*	*berdin* (adj.)	*bardint*
	izutu 'frighten'	*izu*	*izu* (noun)	*izit*
	bihurtu 'give back'	*bihur*	*bihur* (adj.)	*bihurt, bühürt*
	sortu 'create'	*sor*	-	*sort*
	irakurtu 'read'	*irakur*	-	*irakurt*
-ki subclass	*ideki* 'open'	=	-	*idek*
	ebaki 'cut'	=	-	*ebak*
	idoki 'take out'	=	-	*idok*
	iradoki 'force out'	=	-	*iradok*
	jaiki 'get up'	=	-	*jaik*
	eraiki 'erect'	=	-	*eraik*
	atxeki 'hold'	=	-	*atxek*
vowel end	*bete* 'fulfill'	=	*bete* (adj.)	*betha*
	busti 'moisture'	=	*busti* (adj.)	*busta*
	gorde 'hide'	=	*gorde* (adj.)	*gorda*
	erre 'burn'	=	*erre* (adj.)	*erra*
	arazi (causative)	*arazi* → *araz*	-	*eraz*

8.3.2. The third non-finite verbal form: the gerund

8.3.2.1. On the Basque gerund

The gerund is the most frequently used non-finite verbal form in present-day Basque, being present in a wide range of nominalized verb phrases with all the functions that nouns can perform. This is more evident in peninsular dialects, where subjunctive periphrases with the auxiliaries **edin* or **ezan* are regularly replaced by nominalized clauses. These nominalized clauses can fulfill every function characteristic of nouns, hence their high level of occurrence in subordination; in fact, it can be argued that the subjunctive conjugation with **edin* or **ezan* auxiliaries is disappearing from spoken peninsular Basque:

(8) *Nahi* *dut* *Miren* *ibil* *dadin*
want AUX(PRS) p.n. walk(RAD) AUX(IRR).PRS
'I want Miren to walk'

(9) *Nahi* *dut* *Miren* *ibil-tze-a*
want AUX(PRS) p.n. walk-NOMNL-ABS.SG
'I want Miren to walk'

The gerund is also present in the imperfective participle, as this form is the historical result of adding the archaic inessive suffix *-n* to the gerund.[12] At this point, I should refer to a terminological difficulty, since in Basque grammar tradition the term *aditz izen* 'verbal noun' is widely used, but could lead to confusion. In effect, what gerund-deriving suffixes actually nominalize is not the verb, but the whole sentence (de Rijk 2006); besides, such a nominalized verbal form can be identical in form to a mere de-verbal noun deriving from the same verb.[13] Hence, I shall prefer the term *gerund*, the form from which the *imperfective participle* will be derived.

Let us now deal with the formation of the Basque gerund, which is obtained by adding a suffix to the verbal radical. Standard Basque —here I emphasize *standard*, since, as explained below, matters are very different in dialectal and historical Basque— has two allomorphs in complementary distribution for the formation of the gerund:[14] *-te* for verbs with radicals ending in *-n* or a sibilant: *egin* 'do' → *egite*, *ikus* 'see' → *ikuste*, *ikuz* 'clean' → *ikuzte*, *jaits* 'get down' → *jaiste*, *urgaitz* 'help' → *urgaizte* (Basque phonological rules block any affricated + plosive association), and *-tze* for the rest: *hel* 'arrive' → *heltze*, *sor* 'born' → *sortze*, *eduki* 'contain' → *edukitze*, *igo* 'go up' → *igotze* and so on.

The distribution of *-te* and *-tze* is very different from western to central or eastern Basque, and it also reflects a diachronic process whereby the co-occurrence of these allomorphs —historically two different suffixes— has been resolved in detriment of the older *-te*. Facts can be summarized as follows (Urgell 2006: 924-6):

a) In central dialects (Guipuscoan, High Navarrese and, to a lesser extent, also Labourdin), *-te* is clearly a fossil which only remains in verbs ending in

12 While the archaic inessive suffix *-n* joins directly to the word, the modern inessive has incorporated the article *-a.* For *etxe* 'house', modern *etxe-an* contrasts with the archaic *etxe-n*, still used in Souletin. The fact that the imperfective participle takes an archaic suffix (*egite-n*) whilst the inessive construction in nominalized clauses also exhibits the article (*egite-a-n*) makes a relative chronology obvious.

13 "Such verbal nouns [i.e. nominalized verbs] retain their verbal nature in full. […] Several dozen common verbs form derived nouns with *-te* or (rarely) *-tze*; most verbs cannot take these suffixes for this purpose. Like all other derived nouns, nouns derived from verbs by means of *-te* or *-tze* lose their verbal properties completely and exhibit the full range of ordinary nominal properties" (Abaitua & Trask 1987: 398).

14 Nowhere in historical Basque has the gerund been obtained by means of only one suffix.

-n and a sibilant (*egin* → *egite*, *ikus(i)* → *ikuste*), but there are plenty of innovating exceptions such as *hasitze* or *hustutze* instead of *haste* ← *has(i)* 'begin' and *huste* ← *hustu* 'empty'.

b) In eastern Basque, *-te* is more frequently used, since it can also be added to old stems ending in lateral consonants: *ibilte* (SB. *ibiltze* ← *ibil(i)* 'walk'); *erorte* (SB. *erortze* ← *eror(i)* 'fall down'), as occurs in Biscayan. This coincidence between peripheral dialects makes the hypothesis that *-te* was once stronger very plausible.

c) In western Basque, the situation is much more complex: in addition to *-te* and *-tze*, the varieties of Biscay and Alava exhibit a luxurious puzzle of co-occurring suffixes to form the gerund, with a very heterogeneous distribution:
   - *-te* is still widely used: as well as with *-n* class verbs —which lose their *-n* when forming the gerund— and sibilant endings, it is regularly added after lateral consonants (*etorr(i)* 'come' → *etorte*; *ibil(i)* → *ibilte*) and even nasals (*ipin(i)* 'put' → *ipinte*).
   - Verbs with participles ending in *-tu* have gerunds ending in *-tute*; in other words, the suffix *-te* is directly added to participles ending in *-tu*, and the same seems to have occurred with borrowed participles ending in *-idu*: *apurtu* 'break' → *apurtute*, *sufridu* 'suffer' → *sufridute*.
   - The suffix *-tze* remains in a few gerunds, above all in those of one-syllable stems: *har(tu)* 'take' → *hartze*.
   - Recently borrowed verbs with participles ending in *-a(d)u* take the suffix *-eta*: *kantau* 'sing' → *kanteta*.
   - *-keta* is used in the Deba Valley (Biscayan of Guipuscoa) when participles ending in *-tu* have two or more syllables: *apur(tu)* 'break' → *apurketa*; *garbi(tu)* 'clean' → *garbiketa*.

Table 8.4. Non-finite verbal forms

	PFV. PTCP.	RADICAL	GERUND	IPFV. PTCP.
-i class	*ikusi*	*ikus*	*ikuste*	*ikusten*
	bidali	*bidal*	*bidaltze*	*bidaltzen*
-n class	*egin*	*egin*	*egite*	*egiten*
-tu class	*hartu*	*har*	*hartze*	*hartzen*
-ki	*eduki*	*eduki*	*edukitze*	*edukitzen*
-o	*igo*	*igo*	*igotze*	*igotzen*
= adjective	*erre*	*erre*	*erretze*	*erretzen*
= adjective	*hil*	*hil*	*hiltze*	*hiltzen*

8.3.2.2. The gerund, suffix by suffix

Urgell's 2006 work is the most relevant contribution to the study of the Basque gerund (note that she prefers the term *verbal noun*). In this paper, the author systematizes what most linguists and philologists thought about the diachronic evolution of the Basque gerund, by means of a thorough and far-reaching search in pre-literary and literary corpora.

8.3.2.2.1. -te

The gerund-forming suffix *-te* has been identified with the suffix *-te* present in *eurite* 'rainy season' (← *euri* 'rain'), *idorte* 'dry season' (← *idor* 'dry') or *gosete* 'famine' (← *gose* 'hunger'), with an original meaning of 'duration' (Trask 1995b: 221) or 'time of' (Urgell 2006: 925). Its first attestation, in the 14th c., is *ozterate* (*TAV*: 51-56), the name of a military tax in the Middle Ages. After comparing archaic texts and the distribution of the suffix in marginal areas, *-te* is thought to have been used with ancient verbs with [*e-STEM-i] structure (*irakurte* ← *irakurr(i)* 'read'), *-n* class verbs (*egite* ← *egin*), vowel-ending verbs (*ozterate*) and *-ki/-gi* verbs (*jalgite* ← *jalgi* 'go out'). Before 1545, there are no attestations of *-te* with loans or with verbs taking the participle *-tu*.

The recessiveness of *-te* throughout historical Basque is clear from the most ancient texts, where *-te* appears in co-occurrence with the central/eastern *-tze* and the western *-eta*, both firstly and mostly attested with loanwords. The beginning of this process cannot be precisely dated, but we can date its completion in the most innovating geographical area:[15] between 1545 and 1645, the number of verbs taking *-te* were reduced to the two possibilities accepted in the modern standard: *-n* class verbs and stems ending in a sibilant.

Even if both western and eastern areas reflect a situation closer to pre-literary Basque, there is an interesting difference between them: in the eastern area, verbs lacking any participle marker take *-tze* when forming the gerund (*hiltze* ← *hil* 'die', *bethatze* ← *betha* 'fill', *ithotze* ← *itho* 'drown'), while in western Basque the suffix is *-te*: *ilte*, *betete*, *itote* (same gloss).

8.3.2.2.2. -tze

According to Urgell (2006: 930), the origin of the suffix *-tze* would be related to *-*(t)zaha*, transcribed as *-zaha* in the medieval place names of Alava (*TAV*: 25-30), and Manterola (2006: 674) has seen in this *-zaha* the trace of the Basque

15 Urgell delimits this area as a triangle between Southern High Navarrese, the Eastern Low Navarrese of Cize and the Labourdin of Donibane Lohizune and Sara (ibid.: 928).

article *-a* (< **-ha*) added to the suffix *-(t)za*, profusely attested in toponyms. Lakarra has remarked to me, however, that, in forms lacking an article, this *-tza* would have been reduced to *-tz* and, consequently, an *e* would have been necessary when adding the inessive, the allative or any other case marker with an initial consonant: *sar-tz-e-n*, *sar-tz-e-ra*. In this case, speakers would have later reanalyzed these forms as *sar-tze-n*, *sar-tze-ra*, etc. Such a reanalysis is a very common phenomenon, also observable in place names such as *Parise*, *Miarritze* or *Atharratze*. Be it as it may, *-tze* emerged in association with *-tu* participles, regardless of whether they were borrowed or not. Then, a massive [V + *-tu*] borrowing would have facilitated the association of this suffix to vowel-ending verbs: *atera* 'go out' → *ateratze.* Broadly speaking, the more ancient and/or western the text, the more unusual *-tze* becomes.

8.3.2.2.3. -eta

The *-eta* used in the formation of the gerund happens to be the suffix *-eta* widely used in place names, and in the origin of the local cases (see §6.3.2). Its entrance in forming the gerund seems to be closely related to the construction of the imperfective participle with an inessive marker (*egin* 'do' → *egite-**n***), and thus *-eta* could be the most recent suffix involved in forming the gerund. This hypothesis is reinforced by the fact that *-eta* is mainly added to (recently) borrowed verbs which have participles ending in *-adu*, *-idu*: *goardadu* 'keep' → *goardaeta*; *prometidu* 'promise' → *prometieta.*

As Gómez (1991: 411) realized, in the Roncalese dialect, the gerund exhibit(ed) both *-eta* and *-ta*, the latter being a simplified form which would also have supplanted the old —and phonologically close— *-te*. After a detailed search of the literary corpus, Urgell describes an increasing evolution of *-eta* as a gerund-forming suffix in western varieties and proposes that this usage would have been an Alavese innovation (ibid.: 932-3). The author comes to the conclusion that such an innovation, successful in western Basque and amenable to generalization, was blocked by the reinforcing of *-tze* as a regular suffix in central and eastern areas.

8.3.2.2.4. -keta

It is widely agreed that the suffixes *-eta* and *-keta* have a common origin, as Michelena observed (1987 [1971b]: 145). On the other hand, we must bear in mind phonological contexts in order to decide which of these forms is basic and which secondary, and the fact that in toponymy *-keta* was practically limited to

the position after the sibilant makes this issue quite clear (*FHV*: 247). As a gerund-forming suffix, the area where *-keta* first appeared has remained more or less the same —excluding most of Alava, where the Basque language has been substituted by Spanish—: the Leintz-Gatzaga Valley in Guipuscoa and Aramaio, its neighboring land in Alava.

Urgell suggests that *-keta* was indeed an Alavese innovation (2006: 934), unveiling the richness of examples and the complementary distribution of these suffixes in Landucci's manuscript (Vitoria-Gasteiz, 1562). Thus, both *-eta* and *-keta* can be considered to be Alavese innovations. In fact, she proposes with clarity that what irradiated from Vitoria-Gasteiz would not at first have been a pair of suffixes, but a mere morpheme *-(k)eta* with two allomorphs whose use depended on the phonological context (ibid.: 934): in archaic Alavese texts, *-eta* is restricted to vowel-ending stems (*suplikaeta* ← *suplikadu* 'implore'; *kantaeta* ← *kantadu* 'sing'), whereas *-keta* can be added to any verb (*bihurketa* ← *bihurtu* 'return'; *itaunketa* ← *itaundu* 'ask'). Therefore, in the western area, *-(k)eta* has struggled to replace *-tze* as the gerund-forming suffix for verbs with participles ending in *-tu* and, since this *-tu* class is by far the largest one in historical Basque, *-tze* was excluded in favor of the more frequent *-eta.*

8.3.2.2.5. The complex suffix -(t)zaite

In Biscayan and Alavese old texts, we can still find an additional gerund-forming suffix, *-(t)zaite*, present in ancient verbs with a CVC root, in forms which are monosyllabic and mostly ending in a sonorant: *sartzaite* ← *sartu* 'enter'; *galtzaite* ← *galdu* 'lose'; *kentzaite* ← *kendu* 'take away'. In some Alavese texts, *-(t)zaite* seems to be the subject of analogical extension, since it is also used with sibilant endings (*autsaite* ← *(h)autsi* 'break') and even polysyllabic verbs (*garbitzaite* ← *garbitu* 'clean').

On the origins of *-(t)zaite*, two main hypotheses have been formulated. Trask (*HB*: 215) saw the suffix *-te* as being added to the gerund of a *-n* class verb exhibiting *i* (i.e. forms like *izai*, *egoi*, as commented in §8.3.1.2.2). On the other hand, Urgell (ibid.: 938) proposes a morph *-tzai-* which, in effect, also appears in the agent-forming suffix *-tzaile*.[16] This *-tzai-* could have been added as a reinforcing element in one-syllable bases with noun-forming suffixes, as some

16 The suffix *-tzaile* is clearly composed of an unknown element (*-tzai-*) and the well-attested suffix *-le*, also and exclusively used in forming the agent. In fact, both *-le* and *-tzaile* were in complementary distribution depending on the participle marker: *-tu* verbs happened to take *-tzaile,* while the rest took *-le,* no longer productive: *kreatzaile* 'creator' ← *kreatu* 'create'; *igorle* 'sender' ← *igorri* 'send'; *ikusle* 'viewer' ← *ikusi* 'see'.

archaic occurrences suggest: *galtzaigarri* 'cause of ruin' ← *galdu* 'lose'; *sarzaikeran* 'entrance' ← *sartu.* This morph *-tzai-* is also present in another composed suffix, the agent-forming *-tzaile*, which occurs in complementary distribution with *-le*: *egile* 'agent' ← *egin*, *hausle* 'transgressor' ← *hautsi* 'break', *irakurtzaile* 'reader' ← *irakurtu* (but synonym *irakurle* ← *irakurri*).

Apart from anything else, this suffix *-(t)zaite* was indeed related to CVC patrimonial roots. Remember that, from the first records available, these verbs formed the gerund with *-tze* in central and eastern areas, while solutions like *(h)artute* and even *(h)artze* would have been subsequently diffused in western dialects. Thus, *-(t)zaite* has been an unsuccessful solution in the history of the Basque gerund.

8.3.3. The prospective participle[17]

The Basque prospective participle is formed by adding a suffix to the perfective participle. This suffix can be *-ko* or *-(r)en*, in a distribution which ranges from the exclusivity of *-ko* in western dialects to the exclusivity of *-(r)en* in some eastern varieties, with different degrees of co-occurrence in Labourdin, both High and Low Navarrese, and Souletin. Thus, the prospective participle of *egin* 'do' is *egingo* or *eginen*, of *ikusi* 'see', *ikusiko* or *ikusiren*, and of *kantatu* 'sing', *kantatuko* or *kantaturen*, always depending on the dialect. According to phonological rules, *-ko* becomes *-go* after *n*, and the genitive marker *-en* takes a euphonic *-r-* when added to a vowel-ending stem. Both future participles are used in Standard Basque, even if the acceptance of *-(r)en* is restricted to *-n* class verbs. The *-ko / -(r)en* distribution occurs in the three main models shown in Table 8.5.

I am tempted to present the geographical locations for these models, but between II and III, the data are somewhat complex. There is no doubt about the position of Biscayan and Guipuscoan in the first model, and this is clear in historical Basque. In addition, virtually all of the other present-day dialects are to be classed in II, whereas only some Souletin varieties can be located in III. But

17 This point deals with a specific way to express futurity: the main form of the future in historical Basque, well attested in all dialects and the only one specialized for this purpose. Following Mounole (2011: 71-77), in the 15th and 16th centuries there are several analytical constructions expressing future: i) [RAD + **edin*, **ezan* + *-te* / *-ke*] periphrasis, which can also fulfill potential values; ii) [PFV PTCP + *egin*] periphrasis, a specific construction of western dialects, where *egin* is used as AV instead of **ezan* in future and potential; iii) [RAD + **iron*]; iv) [PROSP PTCP + *izan*, **edun*]; v) [IPFV PTCP + *izan*, **edun* + *-te* / *-ke*]. In §9.4.1.3 and §9.4.2, reference will be made to the *-te* / *-ke* suffix.

this distribution has changed over the last centuries, since we have evidence of a recessive evolution of *-(r)en* in records.

Table 8.5. -ko / -(r)en *distribution*

	-ko	*-(r)en*
I (western)	all verbs	-
II (central)	all except *-n* class	*-n* class
III (eastern)	-	all verbs

As stated above, in general, the further east we go, the more range the suffix *-(r)en* covers, and now it can be added that this eastern area has undergone a significant withdrawal throughout the last few centuries. For instance, Dechepare (LN, 1545) uses *-ko* only five times in the 34 occurrences of verbs which could take this suffix. An even more revealing example is Leiçarraga[18] (eastern Labourd, 1571): out of one thousand instances of prospective participles, excluding those of *-n* class verbs, only one is formed by means of *-ko*: *ethorriko* (4 occurrences). In the oldest texts in Souletin, *-ko* is almost non-existent,[19] but at a given moment it shows up and becomes normal, always co-occurring with *-(r)en.*

In the 18th century, the eastern end of the geographical area of the Basque language, represented by the Souletin-speaking village of Esquiule, located in Bearn, offers three relevant texts: two printed books (1758 and 1780) and a popular drama copied down in 1750.[20] Data are as follows: The 1750 drama has 7

18 Among Leiçarraga's merits, we cannot forget his attempt, the first in the history of Basque, to create a standard variety for continental speakers. For this purpose, he chose most of his assistants from Soule. As far as I am concerned, however, the primacy of *-(r)en* in Leiçarraga's works is to be explained by bearing in mind the former extension and strength of this suffix, and not —at least not exclusively— because of his deliberated linguistic choices. Coherently with this argument, *Kadet eta Bettiriño* —another eastern Labourdin text dated ca. 1750— shows a balanced co-occurrence of *-ko* and *-(r)en* in the classes of verbs where *-ko* can be used.

19 Mounole (2011: 74) excludes the prospective *-ko* in Souletin after the single example found in Oihenart's proverbs, but we can offer very different data. In addition to the texts of Esquiule mentioned in this section, Jean de Tartas (17th century) exhibits a distribution which is 90% favorable to *-ko*, even if he represents a marked exception in Old Souletin, since this author consciously adopted some features of other dialects in order to get closer to non-Souletin readers, i.e., in his historical context, Labourdin and Low Navarrese readers.

20 The two printed books are *Sacramentu-Saintiaren Aurhidegouaco maniac eta chediac. Esquiulaco parropia eliçan, Andredonamariaren Conceptionecouan, eçaria…* (Pau, 1758), and *Andere Dona Maria, scapulariocouaren confrarioco bulla, decreta, statutac*

samples of *-ko*, all of which are formed with *-tu* verbs except for the astonishing *edanco* ← *edan* 'drink'; in the 1758 text, all occurrences of the prospective participle take *-(r)en,* whereas in the 1788 text, futures formed with *-ko* represent about a third. Even if it may be somewhat anecdotal evidence, certain eastern varieties of Souletin have developed a pleonastic innovation: *jin* 'come' → *jinen* → *jin**enko***; *egin* 'do' → *eginen* → *egin**enko***. We must note that it is precisely in Soule where other periphrases involved in the formation of the future —and more specifically by means of the suffix *-te* / *-ke*— still remain in use (see at the beginning of this subsection §8.3.3, in footnote). In order to explain the expansion of *-ko* in eastern Basque, a hypothesis can be formulated by the conjunction of some of these factors:

a) The fact that eastern dialects —which generally prefer *-(r)en* as the prospective participle marker— add *-ko* to the non-verbal element of verbal locutions when expressing futurity[21] (Mounole 2011: 77).
b) The fact that the areas where the *-(r)en* destinative and the *-(r)en* prospective participle have been attested are coincident, which impels us to consider that the quasi-extinction of the first has caused the withdrawal of the second (see below, at the end of this Subsection §8.3.3, in footnote).
c) Together with a) and b), the well-known principle of economy in language.
d) The contact with speakers of other varieties and/or some writers' choices.

This said, the *-ko / -(r)en* distribution can be represented in a quadripartite schema in which the mixed model corresponds to a geographical area not yet well-delimited, but historically withdrawing to the east, and nowadays circumscribed to somewhere in the province of Soule. This is shown in Table 8.6.

Table 8.6. Models for the prospective participle

	-tu class	*-i* class	*-Ø* class[22]	*-n* class
I	*sortuko*	*igorriko*	*-ko*	*izango*
II	*sortuko*	*igorriko*	*-ko*	*izanen*
Mixed	*sortuko / sorturen*	*igorriko / igorriren*	*-ko / -(r)en*	*izanen*
III	*sorturen*	*igorriren*	*-(r)en*	*izanen*

eta maniac edo chediac… (Pau, 1780). The manuscript is *Sainte Elisabeth de Portugal*, the oldest preserved pastoral.

21 The verb of a locution can express the future by taking the suffix *-te* / *-ke* or by being analytically conjugated; additionally, the prospective suffix can be directly added to the non-verbal element of the locution.

22 Here, for the sake of clarity and concision, I shall not provide instances of every subclass in the zero class of verbs, since all of these behave in the same way in a given model.

With respect to their origins, both prospective participle-forming suffixes are transparent. As Trask remarks (1995b: 219), after its function of deriving participles passed from the ancient *-i* to the Latin-origin *-tu*, the perfective participle began to be used as a base for the construction of the new future participle in finite verb forms. In order to do so, the different dialects made different choices between the two genitive markers. Following Trask, "it is generally believed that this *-en* is in origin the mere genitive case suffix *-en*, and that this *-ko* is the same morph as the familiar relational affix occurring in *etxeko andrea* 'housewife'". In other words, every dialect made a choice between the possessive genitive *-(a)(r)en* and the local genitive *-ko*.[23] Michelena (1987 [1981]: 48), relating the origin of Basque periphrastic constructions to those of neighboring Romance languages, saw the Basque future participle as a gloss on the medieval Romance forms like *he de ver* (lit. 'I have of see') and *es de venir* (lit. '(s)he is of come').[24] The following section returns to this issue.

8.4. The Basque periphrastic verb: History and Pre-History

At the beginning of this chapter, we have briefly seen how the Basque periphrastic verb is formed by means of one of the non-finite forms —radical, participle or gerund— as MV, together with an AV. Now we shall explore how those forms can be combined in order to express TAM categories (TENSE, ASPECT and MOOD). Our second aim will be to briefly explain the diachronic evolution of Basque periphrases.

8.4.1. Combining non-finite forms with auxiliaries: TAM categories[25]

If a synthetic verbal form contains both semantic and grammatical information in a single word, in an analytic verb such information is distributed between the MV (semantic content and aspectual value) and the AV (argumental, temporal and

23 This choice is also observable when expressing the destinative: apart from the standard *-(a)(r)entzat* 'for' —formed on the genitive marker *-(a)(r)en*— most varieties attest different solutions involving *-ko*: Biscayan *-(a)(r)entzako*, *-(a)(r)endako* and eastern *-(a)(r)endako*, *-(a)(r)entako,* besides the archaic use of the possessive genitive marker with a destinative meaning: *jente* (people) *praube**n*** (poor.GEN.PL), 'for the poor people'. Oyharçabal has remarked to me that the parallel choice in destinative and prospective morphs is a sign of the connection between these grammatical features.

24 That is, in effect, the beginning of the modern Spanish future inflections: *ver he* > *veré*, *ver has* > *verás*, and so on.

25 A detailed description of present-day Basque TAM systems is given in Oyharçabal (2003).

modal information). For the verbs which are capable of synthetic inflection, compact forms are compatible with all the modal values, even if their use of synthetic forms in their full range of possibilities —subjunctive and potential, for example— is nowadays a marked choice which is somewhat archaic; in this respect, periphrastic conjugation is preferred to express every value in a more regular way. Nevertheless, some differences in value remain perceptible between synthetic forms and their composed equivalents, so that in both the imperfective present and past tenses, analytic forms express a habitual meaning, whereas their synthetic pairs have a progressive sense.[26]

Table 8.7. Basque periphrastic conjugation (transitive verbs)

Present	Imperfective	*ikusten du*	*dakus**	'he sees it'	Indicative moods: **edun* AV
	Perfective	*ikusi du*		'he has seen it'	
	Prospective	*ikusiko du*		'he will see it'	
Past	Imperfective	*ikusten zuen*	*zekusan**	'he saw it'	
	Perfective	*ikusi zuen*		'he saw it'	
	Prospective	*ikusiko zuen*		'he would see it'	
Hyp. conditional	Past: protasis	*ikusi balu*		'if he had seen it'	
	Past: apodosis	*ikusiko zu(ke)en*		'he would have seen it'	
	Pres/Fut: prot.	*ikusten/ikusiko balu*	*balekus*	'if he saw it'	
	Pres/Fut: apod.	*ikusiko luke*		'he would see it'	
	Past subj: prot.	*ikus baleza*		'if he saw it'	Non-indicative moods: **ezan* AV
	Past subj: apod.	*ikus lezake*	*lekuske*	'he would see it'	
Subj.	Present	*ikus dezan*		'so that he may see it'	
	Past	*ikus zezan*		'so that he might see it'	
	Hypothetical	*ikus lezan*		'were he to bring it'	
Subj. pot.	Present	*ikus dezake*	*dakuske*	'he can see it'	
	Past	*ikus zezakeen*	*zekuskeen*	'he could see it'	
	Hypothetical	*ikus lezake*	*lekuske*	'he would be able to see it'	
Imperative		*ikus ezazu*		'see it!'	
Jussive		*ikus beza* (*ikus dezala*)	*bekus* (*dakusala*)	'let him see it!'	

[Adapted from Oyharçabal (2003: 250-1)] *Synthetic forms marked with an asterisk have a different meaning to that of their periphrastic equivalents.

26 However, these aspectual nuances are less and less respected in spoken Basque.

Table 8.8. Basque periphrastic conjugation (intransitive verbs)

Present	Imperfective	*ibiltzen da*	*dabil**	'he walks'	Indicative moods: *izan* AV
	Perfective	*ibili da*		'he has walked'	
	Prospective	*ibiliko da*		'he will walk'	
Past	Imperfective	*ibiltzen zen*	*zebilen**	'he walked'	
	Perfective	*ibili zen*		'he walked'	
	Prospective	*ibiliko zen*		'he would walk'	
Hyp.conditional	Past: protasis	*ibili balitz*		'if he had walked'	
	Past: apodosis	*ibiliko z(ateke)en*		'he would have walked'	
	Pres/Fut: prot.	*ibiltzen/ibiliko balitz*	*balebil*	'if he walked'	
	Pres/Fut: apod.	*ibiliko litzateke*		'he would walk'	
	Past subj: prot.	*ibil baledi*		'if he walked'	Non-indicative moods: **edin* AV
	Past subj: apod.	*ibil liteke*	*lebilke*	'he would walk'	
Subj.	Present	*ibil dadin*		'so that he may walk'	
	Past	*ibil zedin*		'so that he might walk'	
	Hypothetical	*ibil ledin*		'were he to walk'	
Subj. pot.	Present	*ibil daiteke*	*dabilke*	'he can walk'	
	Past	*ibil zitekeen*	*zebilkeen*	'he could walk'	
	Hypothetical	*ibil liteke*	*lebilke*	'he would be able to walk'	
Imperative		*ibil zaitez*		'walk!'	
Jussive		*ibil bedi* (*ibil dadila*)	*bebil* (*dabilela*)	'let him walk!'	

[Adapted from Oyharçabal (2003: 250-1)] *Synthetic forms marked with an asterisk have a different meaning to that of their periphrastic equivalents.

In Tables 8.7 and 8.8, I present a general schema of the TAM combinations in Standard Basque,[27] first for the transitive verb *ikusi* 'see' (Table 8.7), and then for the intransitive *ibili* 'walk' (Table 8.8). Note the defectiveness of synthetic forms, as well as the clear division of periphrastic forms in accordance with the auxiliaries for indicative or non-indicative moods.

Furthermore, eastern Basque kept the suffix *-ke* —with its allomorph *-te(ke)* in the (monopersonal) verbs **edin* eta *izan*— for the inflected forms in indicative tense/aspects, carrying a sense of uncertainty, probability or futurity. The usages shown in Table 8.9 are those accepted in the most elevated style in Standard Basque.

27 Some periphrases which are well-attested in historical Basque are excluded, since they remain entirely archaic. Such is the case with the subjunctive-present-based protasis (*ikus badeza*), the so-called aorist (*ikus zezan* 'he saw it', in contrast with its modern meaning of subjunctive) and the handful of combinations involving an AV which are no longer productive, apart from non-standard or markedly dialectal constructions.

Table 8.9. Periphrasis with -(te)ke *suffixed* AV

		TRANSITIVE	INSTRANSITIVE
Present indicative	IPFV	*ikusten duke*	*ibiltzen dateke*
	PFV	*ikusi duke*	*ibili dateke*
	PROSP	*ikusiko duke*	*ibiliko dateke*
Past indicative	IPFV	*ikusten zukeen*	*ibiltzen zatekeen*
	PFV	*ikusi zukeen*	*ibili zatekeen*
	PROSP	*ikusiko zukeen*	*ibiliko zatekeen*

8.4.2. Addition of further elements in periphrastic conjugation

Apart from the canonical [MV + AV] periphrasis, analytic conjugation can incorporate a third element to this basic structure, the most common being the particle *ari* for expressing the progressive aspect, and the participles *izan* or *ukan* to reinforce the perfective aspect, in addition to the eventual use of modal particles like *ohi* (habit) *ote* (uncertainty) or *omen* (to express second-hand information). I shall leave these modal particles aside to focus on the functioning of periphrastic verbs.

8.4.2.1. Remoteness in present perfect and past

The reinforcement of the perfective aspect by introducing a second participle gives what Lafon calls *formes surcomposées*, expressing a sort of remoteness. This second participle can be *izan* 'be' for all verbs or, in eastern dialects, *izan* in concurrence with *uk(h)an* 'have' for transitive verbs:

(10) *Judas* *etorr-i* *izan* *zen*
p.n. come-PFV be(PFV) AUX
'Judas came'

(11) *Guregatik* *bekatu* *egin* *ukan* *du*
we.MOT sin do(PFV) have(PFV) AUX
'He sinned for us'

8.4.2.2. Marking the verbs as a sentence focus

When the verb becomes the focus of an affirmative sentence, western speakers tend to insert *egin* 'do' between MV and AV (13). Such a usage of *egin* seems to be a modern innovation:

(12) *Eneko* *birritan* *jauz-i* *da*
p.n. twice fall-PFV AUX(PRS)
'Eneko has fallen down twice'

(13) *Eneko* *jauz-i* *egin* *da*
p.n. fall-PFV do(PFV) AUX(PRS)
'Eneko has fallen down'

8.4.2.3. Expressing progressive aspect: *ari*, *ibili*, *egon*, *jardun*

As seen above, the present and past of synthetic forms have a meaning of continuous aspect, while the analytic forms express a sense of frequency:

(14) *Ane* *eskola-ra* *joa-ten* *da* *egunero*
p.n. (the) school-ALL go-IPFV AUX every.day
'Ane goes to school every day'

(15) *Ane* *eskola-ra* *d-oa*
p.n. school-ALL 3SG-go
'Ane is going to school'

As there is no such synthetic conjugation for the vast majority of verbs, most of the dialects (i.e. all of them except Biscayan) mark the progressive aspect by means of the particle *ari*, in a typical inessive construction with *izan* 'be':[28]

(16) *Madalena-k* *egunkari-a* *irakur-tzen* *du*
p.n.-ERG newspaper-ART read-IPFV AUX
'Madalena reads the newspaper'

28 Since the *ari* periphrasis is in origin a mere locative construction, its use blocks any expression of morphological ergativity and, consequently, all verbs are treated as intransitive, taking an intransitive AV. Nonetheless, some varieties also mark the transitive / intransitive opposition with *ari*, in a clear innovation. As Mounole reminds us (2008: 589), many languages show a semantic connection between the locative and progressiveness. With regards to the etymology of *ari*, Aldai (2003) links it to the ancient form of the verb *hartu* 'take' ← *hari* (cf. Dechepare's pleonastic *haritu*); grammaticalization explains the loss of the *h-*.

(17) *Madalena egunkari-a irakur-tzen ari da*
p.n.(ABS) newspaper-ART read-IPFV 'ari' AUX
'Madalena is reading the newspaper'

Instead of *ari*, Biscayan has its own way of expressing progressive aspect, making use of the verbs *ibili* 'walk', *egon* 'be, stay' or *jardun* 'be engaged on' as auxiliaries:

(18) *Jon egunkari-a irakur-ten da-go*
p.n. newspaper-ART read-IPFV 3SG.ABS-be
'Jon is reading the newspaper'

(19) *Mikel etxe-a garbi-ketan da-bil*
p.n. house-ART clean-IPFV 3SG.ABS-walk
'Mikel is cleaning the house'

(20) *Karmele-k ardo-a egi-ten dihardu-Ø*
p.n.-ERG wine-ART do-IPFV be.engaged-3SG.ERG
'Karmele is making wine'

8.4.2.4. Further auxiliaries

Without trying to exhaust all the possibilities, we can now consider the use of other autonomous verbs as AV, as used in present-day Biscayan. Indeed, from 1545, both *joan* 'go' and *eroan* / *eraman* 'take' are well attested as AV with a habitual meaning for intransitive and transitive verbs, respectively. Such a usage, formerly common in Basque, has survived in Biscayan until the 19th century. These two examples are taken from Dechepare:

(21) *Bihotz-a d-oa-t ebaki* (Dch. VI, 6)
Heart-ART 3SG.ABS-go-1SG.DAT break(PFV)
'My heart gets broken'

(22) *Mundu hon-ek anhiz jende engana-tu darama.*
world this-ERG a.lot people confound-PFV take.3SG.ERG
'This world confounds a lot of people' (Dch. II, 74)

8.4.3. Historical changes in the TAM system

Since the beginning of the literary period, the Basque verbal system has undergone remarkable changes. Always with regard to auxiliaries, we can see that in 16th century texts —and even later— many instances of periphrasis in the past tense involving auxiliaries **edin* / **ezan* actually have an indicative meaning, unlike in present-day Basque:

(23)

Jaigi	*zidi*	*nagi-a*	
stand.up(RAD)	AUX(*EDIN)	lazy-ART	
erra	*zizan*	*uri-a*	(RS, 294)
burn(RAD)	AUX(*EZAN)	town-ART	

'The lazy (person) stood up and set fire to the town'

(24)

Abraham-ek	*enjendra*	*zezan*	*Isaak*	(Lç., *Math* I, 2)
p.n.-ERG	engender(RAD)	AUX(*EZAN)	p.n.	

'Abraham engendered Isaac'

The relevance of these "anomalous" constructions of the past tense, the so-called aorist, is highlighted by the fact that Lakarra chose this feature as a linguistic criterion to establish the limit between archaic and old stages in his periodization of the history of Basque (1997: 511-7). Furthermore, the changes noticed in the **edin* and **ezan* periphrases encouraged Mounole to undertake an exhaustive analysis of all the types of analytical forms in historical Basque (2006 and subsequent works). In view of the fact that different types of periphrasis may have different origins, she has researched their distinct grammaticalization processes and diachronic evolution, proposing a relative chronology. These are the structures of Basque periphrasis (samples are given in Tables 8.7 and 8.8, as well as in the examples given throughout this chapter):

- Type 1: RAD + aux. **edin*, **ezan*, **iron* (western equivalent: PFV PTCP + *egin*).[29] Examples (2), (5), (6), (7).
- Type 2: PFV PTCP + aux. *izan*, **edun*. Ex. (4), (12).
- Type 3: IPV PTCP + aux. *izan*, **edun*. Ex. (1), (3), (14).

Mounole describes a general reorganization of the TAM system in Basque, which occurred from the 15th to the 18th centuries (2011: 316):

29 Western periphrasis [PTCP + *egin*] would have more recent origins than the other structures in this group. Mounole dates its emergence to the "Old Common Western Basque", after the dialectal split (2007: 15-8).

a) The 3rd type of periphrasis becomes the main form of the present, to the detriment of synthetic forms, which become used increasingly less. There are traces of the ancient present in 15th and 16th century texts: periphrasis of the 1st type in the present and the [PFV PTCP + *joan / eroan*] periphrasis would remain only in secondary usages (non-indicative and habitualness, respectively), which were acquired as a new form of the progressive became more widespread.
b) Periphrasis of the 2nd type develop narrative values, replacing the ancient 1st type, which would specialize in non-indicative usages. Synthetic forms of the past also lose their narrative value to be confined to the imperfective past.
c) A similar replacement takes place in conditional and potential moods, whereby both synthetic forms and periphrasis of 1st type withdraw in favor of the new periphrasis involving *izan*, **edun*.
d) The periphrasis [PROSP PTCP + *izan*, **edun*] becomes the main form of the future in all dialects, and the only one in peninsular Basque.
e) As a result of these processes, the subjunctive emerges as a morphologically marked mood, since **edin*, **ezan* and *egin* auxiliaries have been expelled from the indicative.

In other words, the set of periphrases based on *izan* / **edun* auxiliaries constitutes the axis of the new TAM system. It must be pointed out that the majority of the changes described above began to take place before the historical period of the language; in fact, in the 15th and 16th centuries plenty of forms are attested in co-occurrence. Thus, in the earliest centuries of its literary period, Basque undergoes a transition from a pre-historical TAM system to the 18th century system, which remains more or less the same in present-day language. The next point deals with the reconstruction of such a pre-historical system.

8.4.4. Origins of the Basque periphrastic verb

At the beginning of the literary period, auxiliary verbs were used in different ways, which offers some key information for the diachronic analysis of Basque periphrases. Concerning the auxiliaries involved in the 1st type of periphrasis, only the verbs **edin* 'become' and *egin* 'do' have simple forms —i.e. synthetic forms which do not occur as the AV of another verb— although the occurrences of **edin* in the present tense are subject to constraints. On the other hand, **ezan* and **iron* lack any simple form, since they always appear combined with a verbal radical or the particles *ahal* (potentiality) or *ezin* (impossibility). From this situation, Mounole (2006: 726) deduces an advanced stage in the grammaticali-

zation processes of these verbs in general, and points out different levels in such processes: **ezan* and **iron* seem to be the most grammaticalized AV, whereas *egin* has not yet lost its syntactic and lexical autonomy.

Moreover, the grammaticalization processes concerning *izan* 'be' and **edun* 'have', used in the 2nd and 3rd types of periphrasis, would be hardly advanced in the 16th century, as they were still used in all tenses and moods. Nowadays, *izan* and **edun* show different levels of grammaticalization depending on the dialect. Thus, in western Basque they no longer occur in simple forms and the verbs *egon* and *eduki* have taken their place; but in eastern dialects, *izan* / **edun* have not lost their autonomous use, while *egon* / *eduki* still keep their original meanings of 'stay' and 'hold', with any replacements not yet having occurred.

With regards to the forms used as MV in the three types of periphrasis, their relative chronology is clear, as has been described in §8.3.1.1. The origin of the gerund —dated by Urgell to the centuries after the dialectal split, i.e. the Middle Ages— is closely related to the emergence of a new periphrasis of the imperfective (3rd type). The periphrasis of the perfective (2nd type, also involving auxiliaries *izan* / **edun*) may be at least[30] contemporary: both gerund and participle are derived from the radical. Finally, it would seem that periphrastic forms —perhaps those of the 1st type— already existed when the periphrasis of the 2nd type entered the language.

Luckily, the first record of the Basque language, consisting of two short comments written *in margine* in the Emilian Glosses (11th century), contains the obviously verbal form *dugu* 'we have it', surprisingly identical to the form in modern central dialects. These *glossae* also exhibit negation in a different order, and what we could identify as two participles,[31] one of which (*ajutu*) is to be classified as a *-tu* verb, and the other (*izioqui*) presumably as a member of the *-ki* subgroup of the zero class of verbs. If such constructions are to be considered

30 In Mounole's words: "elles doivent au minimum être contemporaines" (ibid.: 733). If not, the 2nd type of periphrasis could be older than the 3rd, since the formation of the perfective participle is a common feature in all dialects.

31 The formal transparency of these medieval glossae contrasts with their semantic obscurity, not yet elucidated (cf. *TAV*: 41-44). We do not know what *ajutu* and *izioqui* actually were, although morphologically they exhibit the same suffixes as historical participles. The question is whether these forms were adjectives, nouns or verbs: the hypothesis of the grammaticalization of verbal periphrases starts from the premise that adjectives / nouns would have been grammaticalized as MV, whereas the ancient MV would have become AV, but in the case of Basque we have no means of precisely dating such processes. In addition, the aspect of both constructions would change depending on the category of *izioki* and *ajutu*: if they were adjectives or nouns the meaning would be imperfective, but if they were participles of an analytical form, they would express perfective aspect.

verbal periphrases, the analytic conjugation would have a proven antiquity of nearly a millennium; it appears, moreover, that its origins would be in PB (considering the periphrasis of the 1st type).

The supposed antiquity of synthetic forms with reference to analytic conjugation is a topic which has been explored throughout the history of Bascology. However, since the proposal of Lakarra (2006b: 300) that, in the past, not all Basque verbs would have been subject to synthetic inflection, the opposite hypothesis is now emerging among scholars. Lakarra describes the panorama in PB: "When Latin and Basque languages came into contact, this presumably had a reduced class of synthetic verbs —with several dozens, maybe one hundred members—, a fossilized or almost unproductive class, as it would remain 1500 years later and, also, surely, in the dawn of Romance languages" [translation mine, MPM] (2006a: 613). Mounole comes to a similar conclusion: "Ne conviendrait-il pas mieux de penser à une classe de verbes simples ou flexionables fermée, en distribution complémentaire avec une classe ouverte de prédicats complexes de type racine verbale + auxiliare?" (2006: 734).

The question is to determine what the supposed MV of those ancient periphrases with AV **edin*, **ezan* (and even now with *izan* & **edun*!) actually were (see above, within this Subsection §8.4.4, in footnote). A high degree of grammaticalization in the periphrasis of the 1st type —and notably earlier than in constructions of the 2nd type— is very plausible. Finally, even in the case that synthetic forms eventually proved to be older than analytical verbs, they would have come from the grammaticalization process of an even older verbal periphrasis.[32] To conclude, it can be argued that the most ancient forms surviving in the historical period are both synthetic verbs and periphrases based on the auxiliaries **edin* and **ezan*.

8.4.5. The influence of Romance models

Michelena (1987 [1981a]: 48) describes a parallelism between Romance —and more generally western European— and Basque periphrastic forms of the perfective and future. He shows good sense in only mentioning periphrases with *izan* / **edun* as AV, since both imperfective periphrases and constructions involving any of the **edin*, **ezan*, **iron* and *egin* auxiliaries escape such Romance-Basque parallelism, as will be explained. Table 8.10 shows a selection of such parallels in some European languages.

32 Here, Trask (1977) saw in the morph *da-* involved in synthetic forms the trace of an ancient copula: *dator* 'he comes' < *da* + *tor* (*tor* being the MV).

Table 8.10. Parallel structures in European languages

	Basque	French	Spanish	German
'he has came'	*etorri da*	*est venu*	*ha venido*	*ist gekommen*
'he has done'	*egin du*	*a fait*	*ha hecho*	*hat gemacht*
'he will come'	*joanen (joango) da* (cf. old Sp. *es de ir*)	*ira*	*irá* (< *ir ha*)	*er wird gehen*
'he will do'	*eginen (egingo)du* (cf. Sp. *ha de hacer*)	*fera*	*hará* (< *ha(ce)r ha*)	*er wird machen*
[Michelena (1987 [1981a]: 48)]				

Embracing all kinds of periphrasis in a *totum*, Haase (1994, apud Mounole 2006: 730) proposes: "the analytical construction is an innovation that came about in a (pre)-Romance language-contact situation, most probably originating from spoken Latin. The reason for such an assumption is the structural similarity with the *habeo-factum* perfect". Mounole (2008: 601) also refutes the hypothesis formulated in Haase (1992), which identifies the Gascon influence as the origin of imperfective participle-based periphrasis in Basque.

For Lakarra (2006a: 613), it is not at all plausible that a Latin-Romance influence should have dramatically reduced a class of synthetic verbs which were formerly productive, or that such an influence caused the replacement of this class by the periphrastic constructions mainly used in the history of Basque. Only the structure [PFV PTCP + *izan*, **edun*] has obvious Latin-Romance models and chronologies, unlike the periphrases [IPFV PTCP + *izan*, **edun*] and [RAD + **edin*, **ezan*]. Moreover, he suggests that before the influence of Latin-Romance, the class of synthetic verbs was in a minority, whereas the ancient periphrases —not necessarily those used in the historical period, nor with a similar distribution or function— enjoyed an absolute majority, as has occurred in more recent times.

In Mounole's words (2006: 727): "Il convient souligner que si l'emploi de la racine verbale comme verbe principal n'est pas connu des langues qui nous entourent, il est le plus commun et étendu parmi les périphrases des langues du monde", and "les deux autres périphrases [i.e. other than [PFV PTCP + *izan*, **edun*]] n'ont pas de cognat parmi les périphrases des langues qui ont été ou sont en contact avec le basque, ce qui signifie qu'elles sont issues de processus d'évolution internes à la langue basque elle-même" (ibid.: 730). Supporting the opinion that the entire Basque periphrastic conjugation is due to Latin-Romance influence would entail some serious problems in chronology (ibid.: 731).

To sum up, then, the trace of a Latin-Romance influence in Basque periphrastic verbs is limited, at most, to the periphrases based on either perfective or prospective participles with *izan* / **edun* as AV (i.e. the most recent construc-

tions in Basque periphrases). Such an influence could have prompted the emergence of parallel structures in Basque, or simply the increase in frequency of a pre-existing periphrasis.

8.5. Conclusion

Basque non-finite verbal forms may be arranged in a chronological sequence starting from the time of Proto-Basque (radical, from which PERFECTIVE PARTICIPLE would be derived) to develop, after the dialectal split (from the 6th century onwards), a PROSPECTIVE PARTICIPLE and a GERUND —the latter with a high degree of geographical variation.

As far as PB reconstruction has reached in its development, there is no sign of a former universal synthetic conjugation in Basque, and the related idea which attempts to explain the whole periphrastic conjugation as a result of Latin-Romance influence has been sturdily refuted. Such an influence could explain only one type of periphrasis (based on both perfective and prospective participles), whereas the rest of them date from PB (periphrases involving the radical) or from the Middle Ages (periphrases involving the gerund). In accordance with the diachronic evolution of Basque, dialectal differences concerning non-finite verbal forms —and hence periphrastic conjugation— have incessantly increased since the dialectal split.

In short, even if our knowledge of non-finite verbal forms has considerably profited from the general improvement of research in Basque historical linguistics over the last few decades, some questions remain unresolved, and the limited records available oblige diachronists to fully exploit the resources at their disposal. In addition, the absence of any genetically related languages makes reconstruction work especially difficult, hence the importance of internal comparison, assisted —as has been well understood in recent years— by typology.

Appendix. Basque Non-Finite Verbal Forms: an Approximate Relative Chronology

?	Lakarra's Old PB	• **e-* prefix, productive • **da-*, **(-)ra-* prefixes, productive
?-1st c. AD	Michelenian PB	• *-i* suffix, productive
1st-4th cc.	Aquitanian	• **e-* prefix fossilized
5th-6th cc.	Common Basque	• [RAD + **edin*, **ezan* & **iron*] periphrasis • Latin-Romance origin *-tu* suffix enters the language • *-tu* suffix spreads to patrimonial words
6th-15th cc.	Medieval Basque	• *-tu* replaces *-i* as productive PFV PTCP morpheme • Prospective participle: *-ko / -(r)en* • [PTCP + *egin*] periphrasis in Western Basque (B, Al, G) • [PFV PTCP + *izan* & **edun*] periphrasis • Historical gerund arises (*-te* suffix dates back to PB) • *-tze* gerund-forming suffix emerges as a central innovation • *-(k)eta* gerund-forming suffix, a western innovation • Periphrasis of imperfective [gerund + inessive]
15th-16th cc.	Archaic Basque	• Reorganization of the TAM system (16th-18th) • Reanalysis of certain [N-V] verbal periphrases as [MV-AV] • In B, *egin* begins to replace **ezan* as transitive AV for marked moods • In L, LN and HN, **ezan* begins to replace **iron* as transitive AV for marked moods
1600-1750	Old Basque	• The so-called aorist disappears • Complementary distribution of *egin* / **ezan* in B, Al, G • Reinforcement of the radical in eastern dialects • Radical disappears in western dialects (from B to G)
1750-1875	1st Modern Basque	• Literary G generalizes **ezan* as transitive AV for marked moods in detriment of *egin*
1875-1960	2nd Modern Basque	• **iron* defective AV withdraws dramatically in the east • **edin* & **ezan* fossilized in B and G
1960 →	Unified Basque	• Generalization of nominalized clauses in detriment of the subjunctive conjugation with AV **edin* & **ezan*

9. Finite Verbal Morphology

Borja Ariztimuño
(University of the Basque Country, UPV/EHU)

9.1. Introduction*

In this chapter, I give an overview of the Basque finite verb, from both a synchronic and a diachronic point of view. Since the Basque language has a rather peculiar verb conjugation (at least from the point of view of the other European languages), I should specify that here I shall analyze on one hand the (somewhat marginal) synthetic conjugation, and on the other hand the finite verbal forms of periphrastic constructions, i.e. auxiliaries (the synthetic origin of which is generally assumed). This chapter, then, is closely related to the previous one (§8. "Non-Finite Verbal Morphology"), to the extent that it is there that the periphrastic constructions are studied, which include the finite verbal forms contained in them (see also *HB*: 218-236, *GB*: 195-323, de Rijk 2008).

9.2. Background and brief description

9.2.1. Previous works

Whenever the Basque language is mentioned, one of the topics most frequently referred to is its verb morphology. This area has been studied by most scholars who have worked on Basque, and quoting Trask we can say, at least from the point of view of the Indo-European languages, that Basque verbal morphology is "remarkably rich" and "has often seemed rather exotic".[1]

* The research presented in this work is funded by a Predoctoral Grant of the Basque Government [BFI 2009-236], and has been carried out within the Training and Research Unit UFI11/14 of the University of the Basque Country (UPV/EHU) and the Consolidated Research Group "Historia de la Lengua Vasca y Lingüística Histórico-Comparada" (HLMV-LHC), Ref. GIC 10/83, IT 486-10, funded by the Basque Government.
I would like to thank the editor and colleague Mikel Martínez-Areta for having given me the opportunity to collaborate in this volume, and for his deep and insightful readings of my paper. I would also like to thank Joseba Lakarra, who read an earlier version, for all the interesting and helpful discussions that I have had with him.

1 Trask wrote a good, brief introduction to this topic in his "Basque Verbal Morphology" (1981). In this article, he gives a concise account of the main features and problems related to both finite and non-finite verb morphology.

Although the oldest accounts of the Basque verb conjugation date back to the 17th century (Oihenart 1992 [1638/1656],[2] Micoleta 1995 [1653]), and multiplied in subsequent centuries (Larramendi 1984 [1729], Harriet 1741,[3] Astarloa 1988 [1803], etc.), it has been claimed (Trask 1981: 285) that the first serious work was *Le Verbe Basque en Tableaux* (Bonaparte 1991 [1869]). Nevertheless, the most important studies on the Basque verbal morphology, thanks to their innovation and depth, include those by Schuchardt (1972 [1893], 1947 [1923]) and, later and even better, Lafon's monumental thesis (1980 [1944]) and other papers (1999 [1954b]) on this topic.

Since then, there has been no other long monograph on the whole verbal morphology, that is, about *le système du verbe basque*, although the production of high-quality studies on concrete issues has provided more data and has improved working approaches to the topic. These include the following: the prolific but scattered contributions by Michelena;[4] Trask's great works (1977, 1981, *HB*); the paper by Gómez & Sainz (1995) on the origin of the finite forms, and Alberdi (1995) on the development of the terms of address;[5] Lakarra's numerous papers on the features of the ancient dialects (1983, 1986, 1996a, among others) and, more recently, on the formal reconstruction of Proto-Basque (1995, 2006b, 2008a, 2011a, etc.); the papers of Aldai (1994, 2000, 2006, 2008) and his doctoral dissertation (2003) on the grammaticalization of the present and past in Basque; and Mounole's PhD dissertation (2011) on the development of verbal periphrases. Finally, Martínez-Areta's paper on the origin of the diathesis (2012) can also be added to this list.

9.2.2. General traits

Among the different characteristics that have attracted the attention of linguists, the following ones can be highlighted:

---

2 This author's views on the different conjugation types (*synthetic* and *periphrastic*) did not have much influence on later works, but began the tradition of a recurring debate in the historiography of the Basque verb: the genuineness or originality of each conjugation type (Gómez 2008b: 332-333).

3 Unfortunately, not one copy of his two works on the Basque verb conjugation (mentioned by Harriet himself) has survived to our day (Gómez 2006: LII, 260-261).

4 For this, we now have the *Complete Works* (Lakarra & Ruiz-Arzalluz (eds.) 2011), which facilitates the documentation work of the researcher.

5 Although we cannot mention the last two papers as much as they deserve, their contribution is worth recalling here, since they were an inspiration at the time of writing this work.

a) Conjugation types: like many languages, Basque has both synthetic (one-word) and periphrastic (more-than-one-word) conjugation, arranged overall in a complementary distribution. However, there is a partition between those verbs which employ both types of conjugation (whose number has been decreasing historically),[6] and those which have only ever had periphrastic conjugation.[7]
   Moreover, the auxiliary system has historically had a twice binary (hence four-part) division: realis/irrealis and transitive/intransitive.
   Since I shall discuss the morphology of finite verbs, I will analyze both the synthetic verbs and the finite verb forms of the periphrastic conjugation.[8]
b) TAM: especially the manner in which the expression of these categories is performed (§9.4) and, in the case of periphrastic conjugation, the number of possible 'lexical v. – auxiliary v.' combinations.[9] In addition, it is well known that the synthetic conjugation of the Basque verb is rather defective both in aspect/tense and in mood.
c) Agreement system: both in terms of its variety, and the way it is encoded on the verb, the agreement system of the language has always attracted curiosity (see §9.3).

Below, some examples are given which illustrate the general morphological structure of the Basque finite verb (see also Figure 9.1):[10]

6 And probably prehistorically, although this certainly does not mean that, at a remote point in time, every verb might have been conjugated synthetically.

7 I leave aside, for the sake of simplicity, verb stems unattested in non-finite and (therefore) non-synthetic forms, such as *-io-* 'tell', *-i(n)-* 'give', *-di-* 'become', or mere auxiliaries such as *-za-*, *-iro-*, etc.

8 That is, the auxiliary verbs; nevertheless, I will treat these as main synthetic verbs, leaving aside the specific meanings which they usually have in combination with non-finite forms (see, however, §9.7 for a general description of the differences between the two conjugation types).

9 For more details about the periphrastic conjugation and, more specifically, the non-finite verb morphology, see §8.3 and §8.4.

10 Throughout this chapter, I will preferentially use the standard variety of the Basque language (*euskara batua*); specifications are given when the examples are from another dialect or from a specific literary source.

(1)	a. *n-a-bil*	b. *g-ø-oa-za-k/n*	c. *ø-da-txe-z-ki-k/n*
	1SG-PRS-come	1PL-PRS-go-PL-2SG.ALF	3-PRS-be.attached-PL-DF-2SG.ALF
	'I'm walking'[11]	'We are going, (wo)man'	'They are attached to thee (M/F)'

(2)	a. *n-a-kar-ø-te*	b. *ø-da-rabil-tza-t*	c. *ø-d-io-ts-o-k/n*
	1SG-PRS-bring-3-PL	3-PRS-know-PL-1SG	3-PRS-tell-DF-3-2SG.ALF
	'They're bringing me'	'I'm using them'	'Thou (M/F) art telling it to him'

(3)	a. *n-en-go-en*	b. *z-ø-e-torr-en*	c. *l-ø-ø-itza-teke*
	1SG-NPRS-stay-PST	PST-3-NPRS-come-PST	HYP-3-NPRS-be-IRR
	'I was (still) at'	'It was attached to it'	'It would be'

(4)	a. *n-in-duka-ø-n*	b. *n-e-ki-ke-ø*	c. *z-ø-ø-i-zu-ø-n*
	1SG-NPRS-hold-3-PST	1SG-NPRS-know-IRR-3	PST-3-NPRS-give-2PL[ALP]-3-PST
	'(S)he had me'	'I would know it'	'(S)he gave/had it to you' ~ '(s)he had it'

The first two sets (1-2) represent the present tense, and the others (3-4) show the main patterns of the past tense, and of the hypothetical one (3c, 4b). The examples with odd numbers are of intransitive verbs (monovalent and bivalent), that is, those lacking an agentive argument, while the even ones are (mono-)transitive or ditransitive. Moreover, example (1b) has an extra agreement marker which corresponds to the addressed person —the so-called allocutive— in the familiar register, indicating the sex of the interlocutor. So does (4c) in the polite form of address, in which no such gender distinction is made.

As can be observed in these examples, TAM and person-number morphemes are attached around the root,[12] subject to certain factors such as the specific tense and the type of arguments conveyed by the verb. Another striking and only partially resolved issue is also revealed here: the close relationship between the past and the hypothetical tenses.

Thus, the Basque finite verb agrees in person and number with (overt as well as omitted) subjects of intransitive clauses (S), direct objects or patients (P) (both corresponding to absolutive NPs), the subjects of transitive clauses or

11 As I shall expound later (§9.4; see specifically §9.4.1.3), the modern sense of the synthetic finite verbs in Basque is a kind of imperfective, most likely to be translated into English with *-ing* forms, simple forms or both of them; thus, whenever it is possible to use simple forms, they should be understood to convey an imperfective meaning.

12 I will treat them as prefixes when they precede the root, and as suffixes when they follow it.

agents (A, ergative), and indirect objects or recipients (R, dative), as well as with the addressed person when required. By way of illustration, among the examples above we find: *n-*, *g-* and *ø-* as S in (1, 3); *-t*, *-ø*, *-k/n* (2), and *n-*, *-ø* (4) as A; *n-*, *ø-* (2, 4), and *-ø* (4) as P; *-k/n* (1c), *-o* (2c), and *-zu* (4c) as R (sometimes preceded by a pre-dative marker called the *dative flag* (DF): *-ts-*, *-ki-*); and, finally, *-k/n* (1b), and, again, *-zu* (4c) as addressees. Similarly, there are different pluralizers depending on the person and the argument affected by them: *-z(a)* in (1b-c) for plural S; its other variant *-tza* (2b) for plural P; and *-te* (2a) for plural A. And, finally, certain other morphemes have to be added to express TAM categories: for the present, *da-* (with its allomorphs *d-*, *a-*, *ø-*); for the non-present, *e(n)-*, *in-* (seldom *ø-*); for the past, *z-* (only 3rd p.) and *-(e)n*; for the hypothetical, *l-* (only 3rd p.); and *-ke* or *-teke* for a kind of non-definiteness (denoting future/doubt/potentiality...), usually treated as a modal marker.

To sum up, the general structure of the finite verb forms can be outlined as illustrated in Figure 9.1, leaving aside the pluralizers and the allocutive agreement (cf. *HB*: 106). The minimal overt structure is $Tense_1$-V (*da-bil* '(s)he is walking') and one of the more complex ones would be A-$Tense_1$-V-DF-R-Mood-$Tense_2$/CPL (*n-e-kar-ki-zu-ke-en* 'I would have brought it to you').

S/P~A + $Tense_1$ + V + DF + R + Mood + A~P + $Tense_2$/CPL

Figure 9.1. General structure of the finite verb form

9.3. Person and number agreement

Generally speaking, the Basque finite verb must obligatorily agree with all the core arguments of the clause (i.e., ergative, absolutive and dative phrases) both in person and number, making up a full three-person/two-number distinction system.

The person agreement morphemes are, as accepted in SB,[13] those shown in Table 9.1. Note that the 3rd p. has almost no proper marker (see below on the R marker *-o*), since it is the unmarked category —the non-person—,[14] but it also

13 For further information about dialectal variation see §9.8.1 and Chapter 2.

14 I use the term *unmarked* in a synchronic sense, i.e., as the result of other factors that may not have had anything to do with such an iconic concept. In line with Haspelmath (2008), I believe that frequency of use may have been greatly relevant for the development of this and other *empty* categories. This reminds us of Schuchardt, when he writes about the "anomalies" (i.e., irregularities) of the Basque language and languages in general: "Aber nicht bloss in Ausnahmsfällen und nicht bloss die absolute, auch die relative Häufigkeit

distinguishes between the singular and plural number with the addition of different pluralizers, depending on various factors (§9.3.2.1).

Table 9.1. Person agreement morphemes of SB. finite verbs

	Prefixes	Suffixes		
1sg	*n-*	*-da-*	>	*-t#*
2sg[15]	*h-*[16]	*-a-/-na-*	>	*-k/-n#*
3sg	*ø-*	*-ø(-)/-o(-)*		
1pl	*g-*	*-gu(-)*		
2pl[17]	*z-*	*-zu(-)*		
3pl	*ø-*	*-ø(-)*		

The prefixes are said to bear the reference of the S (***n**abil* '**I** am walking'), or the P of transitive clauses (***n**akar* '(s)he is taking **me**'). However, there is a TAM/person split which leads these prefixes to also be co-referential with the A in the past tense (***n**uen* '**I** had it', ***n**ekarren* '**I** brought it') and in forms from the so-called *hypothetical tense* (***n**uke* '**I** would have it', *ba-**n**ekar* 'if **I** brought it'). This phenomenon occurs only when the P is 3rd person; otherwise, the prefixes encode the patient, also in non-present tenses (***n**induzun* 'you had **me**', ***n**indekarke* '(s)he would take **me**').[18]

---

und Gëlaufigkeit der Formen muss berücksichtigt werden, was meines Wissens in der Sprachforschung bisher nicht, wenigstens nicht auf systematische Weise geschehen ist" (Schuchardt 1972 [1893: 53]).

15 The morpheme alternation in the suffixes of the 2nd p. sg. depends on the addressee's gender, being *-a-* or *-k* for the masculine, and *-na-* or *-n* for the feminine.

16 This is the standard (written) variant, adopted from the eastern dialects. In central and western ones, the phoneme /h/ no longer exists, and in some Navarrese varieties, the pronoun *hi* 'thou' and its agreement *h-* are respectively pronounced *yi* [ji] and *y-* [j-]. The loss of the aspiration in some dialects has also led to some changes in the 2nd p. sg. agreement; e.g. an old *(h)uan* 'thou hadst' has been reshaped as *u**k**an/u**n**an* in some varieties where the *h-* has been lost, inserting the suffixes *-k* and *-n*, cf. *du**k**/du**n*** 'thou (M/F) hast'.

17 Inasmuch as the modern (polite) 2nd singular person (*zu*) was historically the 2nd plural (like Eng. *you*, Sp. *vos*), it employs on the verb the same number morphology as the other plurals, and the current 2nd person plural (*zuek*, agreement suffix *-zue(-)*) has a pleonastic characterization of its plurality. Hence, in this paper *hi* 'thou' will be the sole 2nd singular, whereas *zu* 'you (sg. < pl.)' will be presented as the morphologically plural partner of that person, along with *zuek* 'you (pl.)'. Likewise, the "affective" variant *xu* will not be systematically specified, since it shows no other difference beyond its phonetic shape.

18 There is also a difference in the allomorph of the non-present marker, depending on the person/argument which the preceding agreement morpheme corresponds to (see §9.4.1.1.2).

The suffixes, on the other hand, have two meanings: they either carry the reference of the A in the present tense (*du**t*** '**I** have it', *dakar**t*** '**I** bring it'), and also in the non-present if P ≠ 3rd p.; or they bear the meaning of an R in any tense (e.g. *dakarki**zut*** 'I am bringing it **to you**', *nekarki**zun*** 'I brought it to **you**'). The issue, however, is somewhat more complex.

With regard to the 1st and 2nd singular persons, these suffixes change their shape depending on what follows them, or, rather, whether anything follows them or not. More precisely, they appear in the *a-grade* form when followed by another element such as an overt agreement marker (in that case, with the obligatory meaning R: *di**da**zu* 'you have it **to me**'), or by a complementizer suffix heading the finite verb (thus conveying either the R agreement, if there is also a non-overt marker (3rd p.), or the A agreement, if there is not: *di**da**-ø-n* 'which (s)he has it **to me**', *du**da**la* 'that **I** have it'); and they are reduced to the *zero-grade* when followed by either nothing (then A), or by only the zero-morpheme of 3rd person (R) (cf. *du**n*** '**thou (F)** hast it', *di-**ø**-**n*** '(s)he has it **to thee (F)**'). Consequently, while in general the order of the affixes will tell us their meaning (as in *di-**zu**-gu* 'we have it **to you**' vs. *di-**gu**-zu* 'you have it **to us**'), sometimes we are able to guess the actual meaning with the sole help of other elements such as the verb root or the dative flag (thus, *-t* = A in *du**t*** 'I **have** it' and in *darama**-t*** 'I carry it', but *-t* = R in *di**t*** '(s)he **has** it **to** me' and in *darama-**ki**-t* '(s)he carries it **to** me').

9.3.1. Person

In this subsection, I will more thoroughly explore some issues concerning the person agreement markers, both with regards to their origin and their relative position on the verb.

Firstly, we should consider the form of the prefixes. Some authors (Lafon 1980 [1944], Trask *HB*, Martínez-Areta 2012) have differentiated two sets of prefixes as belonging to the aforementioned TAM split. That is, along with the forms in Table 9.1 (which would be present tense agreement markers),[19] they include the forms *nen-*, *hen-*, *gen-*, and *zen-* for the past tense. Leaving aside their respective considerations, I believe that the most useful and economical way of analyzing them, at least from a strictly synchronic point of view, is *X-en-*, parallel to the present tense *X-a-*.

In this way, they seem to actually come from their respective free pronoun (Lafon 1980 [1944]: 374): *n-* < *ni* [1sg], *h-* < *hi* [2sg], *g-* < *gu* [1pl], and *z-* < *zu*

19 And some others also take *na-* as a whole (cf. Schuchardt 1972 [1893]).

[2pl]. The same is entirely true for the suffixes of the plural persons: *-gu* and *-zu* (Lafon 1980 [1944]: 389-390). But those of the singular ones (*-t*, *-da-* and *-k/-n*, *-a-/-na-*) are without any doubt unrelated to the pronouns themselves (ibid.), although some linguists have seen in this alternation (i.e., *n-* ~ *-t*, etc.) the reflex of an ancient stem variation in the pronouns (Trask 1981: 296).[20]

On the one hand, the allomorphs *-da-* and *-t* for the 1st person singular are related to each other by means of a prehistorical change[21] which led to the loss of the final *-a#* of various suffixes and even lexemes (cf. *bart* 'last night' < *bar-da*, retained in this form only in eastern dialects). Thus, once there were pairs like **duda* 'I have it' : *dudan* 'which I have' or **dizuda* 'I have it to you' : *didazu* 'you have it to me'. Then, the first ones underwent final vowel truncation, and became *dut* and *dizut*, whereas the other ones have wholly maintained their phonological structure.

The morpheme itself seems to have partners in other areas of the language, both in the noun phrase and in the verb morphology (Lakarra 2006b, 2008a). In the latter, we should mention the present prefix *da-* (see §9.4 below), and in the case of the NPs, we could mention the *-t* extension in the allative of the eastern Basque varieties (*-rat* vs. western *-ra*), which presumably comes from an older **-da*, along with the *-da-* element found in some temporal adverbs such as *gazte-da-n-ik* [youth-*da*-INES-PTV] 'since (sb.'s) youth', or *gaurdaino* [today-*da*-TERM] 'until today', from which we can deduce a vague meaning of location. Nonetheless, how it was transferred from the NP to the VP (or *vice versa*) remains an unresolved question, as well as why or how it specialized as the agreement of the 1st person singular.[22]

---

20 As I have mentioned above, Martínez-Areta (2012) also proposes a morphological alternation (of the former free pronouns), but in this case, in order to explain the prefix alternation related to tense distinction: present *n(a)-*, *h(a)*, *g(a)-*, *z(a)-* vs. non-present *nen-*, *hen-*, *gen-*, *zen-*. However, I shall analyze this *-en-* as a TAM marker, so that we do not have to explain such a stem alternation (see §9.4).

21 That is, prior to the first written attestations. However, this is not entirely true, since there are some old instances of the ablative (*-tika*, modern *-tik*) and the partitive (*-(r)ika*, modern *-(r)ik*) suffixes which show us that previous stage (see §6.4.2.3).

22 Lakarra (2008a: 603, fn.104), quoting Mounole, recalls that there are some languages with certain locative morphemes deployed as personal agreement markers. For example, according to Zavala (2006), in Olutec (a Mixean language) a 1st p. pl. marker has been grammaticalized from an 'exist' verb (*ʔit*). In the same way, the Basque agreement suffix *-da* could be related (as mentioned above) to the present prefix, which is supposed to come from a grammaticalization of the root of the reconstructed posture verb 'sit' (**dar*). Moreover, that verb is possibly the source of the 3rd p. sg. copula *da*, that is, the Basque verb which is closest to the idea of 'exist' in the language. Therefore, this could be another example of the relationship between 1st person agreement and the 'exist' verb, but

Besides the 2nd sg. pronoun *hi* (> prefix *h-*), we find the suffixes *-a-/-k* for a masculine listener and *-na-/-n* for a feminine one. The first pair definitely comes from the reconstructed form **ga*. When **-ga* was not in final position, it was subject to the usual deletion of the voiced stop between vowels (**du-ga-la* > *duala* 'that thou (M) hast it'), but in the absolute final position it underwent the aforementioned *-a#* deletion and the subsequent devoicing of the obstruent: **dug(a)#* > *duk* 'thou (M) hast it'.[23]

Moreover, it seems to be somewhat obscurely related to the ergative marker *-k* (*ume-a-k* 'the child (A)'), or, preferably, with its possible source: the marker *-ga-* (Lakarra 2005, 2006b). This morpheme remains in the local cases of the animate NP case paradigm (*etxe-a-n* 'in the house' vs. *ume-a-**ga**-n* 'in the child') (see §6.4.2.1). In addition, this morpheme might have come from an old locative suffix (cf. the surname *Ares-pacoch-a-ga* [oak-single-DET-LCV] 'the place of the single oak'), which also led to an abundantial interpretation, probably through ambiguous place names such as *Harri-a-ga* 'the place of the stone(s)' (*harri* 'stone'), or *Pago-a-ga* 'the place of the beech tree(s)' (*pago* 'beech'), until it became a plural suffix of NPs, e.g. *harri-a-k* 'the stones', *pago-a-k* 'the beech trees').[24]

As already mentioned, there is also a morpheme *-na-* ~ *-n*[25] for the 2nd p. sg. feminine. These variants are also easily related to each other by the *-a#* truncation rule. Some authors have considered the feminine agreement marker to be secondary with regard to the masculine one, i.e., that it is formed on an allegedly gender-unmarked **-ga* plus another unknown feminine marker (Jacobsen 1975: 23, apud *HB*: 296, suggested that it should be reconstructed as **-naga*).[26] Be it

---

in the singular, unlike in the Olutec language. Interestingly, on the other hand, the relative sg. form of the Basque copula (*den*) has also developed a plural-like meaning, but lexicalized as a quantifier: *dena* 'all' (< 'which is').

23 In western varieties, the allomorph *-a* (< **-ga* / V_#) can be found in some forms, but this might be due to an analogical leveling with other forms of the paradigm (*deua* '(s)he has it to thee (M)', *deuat* 'I...', *deuagu* 'we...').

24 A more thorough discussion of this issue can be found in §6.3.1.

25 In western varieties, it can also retain the vowel in final position (Aurrekoetxea 1994: 1007).

26 To accept this reconstruction, we would have to assume an early loss of the intervocalic *-g-* followed by a reduction of the resulting double vowel. This is not inconceivable, but along with the **-ga(-)* > *-a-*, *-k#*, we would expect **-naga(-)* > *-na(a)-*, ***-nak#*. Furthermore, the absence of intervocalic *-n-* > *-ø-* (*FHV*: 299 ff.) in forms such as *dunala* 'that thou (F) hast it' could support the hypothesis of a secondary (hence, late) origin of the suffix for 2nd p. sg. feminine, although it may be understood as maintained after the former consonantal ending of that root (**da-du**__**n-n**__**a-la*), if it is not an instance of analog-

as it may, the gender distinction itself seems to be a relatively late development, since it is not present in the pronoun itself (*hi*), nor in the presumably older agreement marker, i.e. the prefix *h-*.

Finally, we should note the lack of overt agreement for the 3rd p. singular. However, for many years, the prefixes *d-*, *z-* (Biscayan *ø-*), *l-* (and imperative *b-*) have been interpreted as 3rd person markers, allegedly coinciding with the TAM division (see §9.4). Regarding the morpheme *-o-* (cf. (2c)) and its variant *-a-*, there have been some attempts to account for its origin. For example, Jacobsen's explanation (1975: 21) is that it comes from a hypothetical element **-do-*, "similar in nature to the first and second person markers" (apud Trask 1981: 296), and also related to the prefix *d-*. But the most successful explanation has been the one defended by Schuchardt. He derived both *-o-* and *-a-* from the demonstrative stems *(h)au*, *(h)or-*, *(h)a(r)-*. However, this could be simply a kind of dative marker, which in some varieties also appears in the plural (along with a pluralizer, see §9.3.2). The lack of any genuine overt agreement for the 3rd person can be understood to be the result of two typologically related factors: the lack of proper pronouns for the 3rd person, and its unmarkedness. However, for better understanding, and because it has had no other function historically, I shall gloss *-o-* as 3rd person.

9.3.2. Number

When we discuss number encoding on the Basque finite verb, there are three factors which must be taken into account: i) the inherent number of an overt agreement marker (thus, only applicable for the 1st and 2nd persons); ii) the explicit marking of that inherent number (again, only with 1st and 2nd p.); and iii) the plurality of the 3rd person. Here are the characteristics of each factor:

a) The inherent number (1st and 2nd p.).

Since the agreement system of the Basque language is largely formed by the grammaticalization of cliticized pronuns, the same distinction of two persons and two numbers of the original set of free pronouns is kept on the verb (see Table 9.1). The shift that has occurred in the former 2nd person plural ('you pl.' > 'you sg. polite') is worth noting. This change began with the free pronoun, and, as in Spanish, a new pronoun was created for the plural with the addition of an element expressing plurality (*zu* → *zuek*; cf. Sp. *vos* → *vosotros*). While the form of the new plural pronoun is well established in all dialects from the beginning of the written tradition, in old texts we can still

ical leveling (cf. *dakarna* 'thou (F) bringst it', in which interestingly the final *-a* does not drop).

find examples of *-zu* as the agreement for the 2nd p. plural, instead of the more recent *-zue* (Laz. ***zuek*** *jakingo do**zu*** 'you may know'; or *ene seme alaba txipi laztantxo**ak** ofizio zitalori eutzi egi**zu*** 'my dear little sons and daughters, give up this vile trade!', in a 17th c. text against witchcraft, cf. *TAV*: 165-167,). Today, this marker (even keeping the pluralizers) is interpreted as singular, unlike e.g. Eng. *you*, but in most dialects it has not entirely replaced the other singular, which is, however, restricted to the most familiar form of address.[27]

b) The explicit marking of the inherent number (1st and 2nd p.).

In almost all of the varieties of the language, and from the earliest attestations, the inherent number is not enough for a finite form to be grammatical when the 1st or the 2nd person is the P or the S of the clause (but it is if they are the A or the R). Compare the following examples:

(5)	a. *ø-d-u-t*	b. *ø-d-u-gu*	c. *n-a-u-ø*	d. *g-a-**it**-u-ø*
	3-PRS-have-1SG	3-PRS-have-1PL	1SG-PRS-have-3	1PL-PRS-PL-have-3
	'I have it'	'We have it'	'It has me'	'It has us'
(6)	a. *ø-za-i-t*	b. *ø-za-i-gu*	c. *n-a-tza-i-o*	d. *g-a-tza-i-**zki**-o*
	3.PRS-be-DF-1SG	3.PRS-be-DF-1PL	1SG-PRS-be-DF-3	1PL-PRS-be-DF-PL-3
	'It is to me'	'It is to us'	'I am to it'	'We are to it'
(7)	a. *ø-d-oa*	b. *ø-d-oa-**z***	c. *n-oa*	d. *g-oa-**z***
	3-PRS-go	3-PRS-go-PL	1SG.PRS-go	1PL.PRS-go-PL
	'It is going'	'They are going'	'I am going'	'We are going'

The four forms of each verb have the same structure (i.e., P-T-V-A, S-T-V-DF-R, S-T-V), but in (5d), (6d), (7b), (7d), there is a pluralizer which (re-) marks the plurality of the P or the S (cf. ***dit**ut* 'I have them', *zai**zki**t* 'they are to me' vs. (5a) and (6a)). Pluralizers are also found in 2nd p. sg. polite forms, because of their origin as plurals (*zai**tu*** '(s)he has you', *zoa**z*** 'you are going'). Nevertheless, in the western dialect there are some local varieties in which certain transitive verbs (as (5d)) lack the pluralizer (e.g. *zau*, *gau* instead of *zaitu*, *gaitu*), mainly when the P is other than the 3rd person (Zuazo 2006: 86), i.e. when the pluralizer would otherwise be pleonastic or synchronically contradictory as is the case with the 2nd p. sg. polite forms. However, although there are a few early attestations of this phenomenon (e.g. Laz. *serbiduko **zaut*** 'I will serve you'), it is not clear whether they are an innovation or an

27 See §9.6.2 for a more detailed description of this register.

archaism. In addition, in these varieties there is an old tendency for the object of bivalent intransitive verbs (6), or the object of ditransitive ones, not to make plural agreement with the 3rd person, (*paga euskun danak* '(s)he paid us all the things' (Zuazo 2006: 86), with *euskun* [(s)he-it-to.us] instead of *euskuzan* [(s)he-**them**-to.us]), but they do agree in common intransitives such as those of (7).

c) The plurality of the 3rd person.

As I have said, the 3rd p. is the non-person, in the sense that it does not have an overt agreement marker (nor a proper personal pronoun). Nevertheless, the Basque finite verb distinguishes between its singular and plural number too, by means of various pluralizers, whether they be S (8a), P (8b), A (8c) or R (8d):

(8) a. *doa-z* 'They are going' b. *dakar-tza* '(S)he is bringing them'
c. *dakar-te* 'They are bringing it' d. *datorki-e* '(S)he is coming to them'

Pluralizers vary depending on different variables: the verb system, the number of arguments indexed and the features of the pluralized argument (person, role). In the following paragraphs, I provide an outline of the main types and their variants, discussing their use and origin:

a) *-tza*: This is the most common S/P-pluralizer. It has several allomorphs, especially in the dialectal varieties, and seems to be related to the abundantial suffix *-tza* of nouns: *elurtza* 'snowfall' (*elur* 'snow'), *hondartza* 'beach' (*hondar* 'sand'). It also resembles the suffix *-tza* for deverbal nouns, which expresses the action or the outcome, that is, which spans the whole process: *aitortza* 'confession' (*aitortu* 'to confess'), *ezagutza* 'knowledge' (*ezagutu* 'to know (about)'). The affricate consonant is the strong variant of the older *-za*, which presumably originated after consonants (cf. (8a) vs. (8b)).

Its standard variants are: *-z*, from the older variant *-za* through the final vowel truncation rule, and *-zki-*, from a reinterpretation of the *-z-ki-* [PL-DF] suffix chain.

The form *-z* has become so widespread in modern Biscayan that it has almost replaced all the other S/P-pluralizers: common *ditu* '(s)he has them' → B. *dituz* → B. *dauz* (cf. B. sg. *dau* '(s)he has it' vs. *du*). The form *-zki* has also spread to other non-dative forms: cf. pairs such as *darama-**ki**-t* '(s)he is bringing it **to** me' : *darama-z-**ki**-t* '(s)he is bringing them to me'; *dit* '(s)he has it to me' (where apparently there is no DF) : *di-z-**ki**-t* '(s)he has them **to** me'; and *dakit* 'I know it' : *daki-zki-t* 'I know them'.

Dialectal variants are the following: *-tz* (from *-tza*, like *-za* > *-z*); and eastern *-(t)zi*, perhaps from *-(t)z-i*, like *-z-ki* but with another variant of the DF.

b) *-de*: This is the sole A/R pluralizer, although it only works with the 3rd and the new 2nd plural persons. And in these cases, in the standard variety, it actually appears either as its strengthened variant *-te* (only 3rd p. A) or as the reduced one *-e*. Both Jacobsen (1975) and Trask (1995b) proposed that the voiced variant was the oldest one, although the former reconstructed it further as **-dee*, and then analyzed it as **-d-ee*, whereby **-d-* would originally be the aforementioned pronoun for the 3rd person and **-ee* the pluralizer itself).

The old form occurs, in the standard variety, in only one verb (*egon* 'stay, be at'), surprisingly pluralizing the S with all persons (*daude*, *gaude*, *zaude* 'they, we, you are' < *dagode*, *ga(*g)ode*, *za(*g)ode* vs. Biscayan *dagoz*, *gagoz*, *zagoz* or bivalent standard forms *dagozkio*, *gagozkio*, *zagozkio* 'they, we, you correspond to it'). Non-standard forms of *izan* 'be' can also contain this suffix, even in the singular, in stylistically marked usage (*garade* 'we are', *nintzaden* 'I was', cf. *gara*, *nintzen*). Finally, we find the old form of the suffix in some local varieties of Biscay, where other dialects use the *-e* or the *-te* variants; that is, with the 3rd p. plural A of the verb **edun* 'have': *daude* 'they have it' (cf. *dau* '(s)he has it') vs. *daue/dabe*, *due*, *dute* of other varieties.

Some attempts have been made to discover its origin by comparing it to the suffix *-te*, which has a kind of abundantial/iterative meaning and is also used to form deverbal action nouns (cf. the above mentioned *-tza*): *elurte* 'snow season/fall', *gerrate* 'war(time)' (*gerra* 'war'), *gosete* 'famine' (*gose* 'hunger'), and *egite* 'deed' (*egin* 'to do'), *esate* 'saying, instance' (*esan* 'to say').

Recently, Lakarra has proposed reconstructing this suffix as a grammaticalization of the Proto-Basque verb root **den* 'finish', adducing some parallels in Mixe-Zoquean languages (Lakarra 2008a).

c) *-it-:* The P pluralizer of a few verbs like **edun* and the auxiliaries **ezan* and eastern **iron*. This is one of the most mysterious pluralizers, both because of its form (apparently VC, whereas the others are CV) and its relative position with respect to the verb root (before it). The latter could mean that it is very ancient, which would be the reason why it is not easy to discern its origin.

Moreover, the form in which we mention it in this work is a mere convention, but it is so pervasive and widespread that it appears, in fact, to be a real form, when actually it is not even a reconstructed one (cf. Heath

1977). Indeed, the *-t-* of *ditu*, etc., could be the devoiced variant of the original **d-* of the root **dun*, presumably resulting from the reduction or the development of a consonant cluster **-C.d-*. Couples like *nau* vs. *gaitu* (cf. (5c-d) above) are misleading in that the first would come from **na-du* and the second from **ga-iC-du*. Therefore, it could be reconstructed as *-CiC-*, since a Proto-Basque morpheme (< root) must have had a consonantal onset too.

To sum up, we can outline the pluralizer system as follows:[28]

1. Subject/Patient pluralizers:
   1.1. *-za* and its allomorphs (*-(t)z(a)*)/derivations (*-zk(i)-/-(t)zi-*).
   1.2. *-de* (only with *egon* 'stay'), and the variant *-te* (with the intransitive auxiliary **edin*).
   1.3. *-it-* (with **edun* 'have', and the transitive auxiliaries **ezan* and **iron*), + *-te* (for the new —pleonastic— 2nd p. pl.).
2. Agent pluralizers:
   2.1. *-te* (in SB only for the 3rd p. pl.).
   2.2. *-e* (in SB only for the new 2nd p. pl.).
3. Recipient pluralizers:
   3.1. *-e* (in SB only for the 3rd pl. and the new 2nd p. pl.).

Here, I should also mention the special case of an irregular formation of the plural form. The verb *etzan* 'to lie (down)', for which the synthetic finite forms for singular persons in the present indicative are *natza* 'I lie', *hatza* 'thou liest', *datza* '(s)he lies', forms its plural with the addition of a *-u-* between the tense vowel and the verb root: *gautza* 'we lie', *zautza* 'you lie', *dautza* 'they lie'. In general, scholars accept Schuchardt's (1907) hypothesis, according to which the *-u-* is the same as that of *gaude* 'we are (staying)', *zaude* 'you...', *daude* 'they...', that is, forms such as *gautza* would originally have been suppletive plural forms, taken from the verb *egon* 'to be at, to stay': < **gagotza* 'we are (staying)' (cf. *gaude* 'id.' < **gagode*, B. *gagoz* 'id.'), and so on.

Now, I would like to try to shed some light on the possible relative chronology of the different pluralizers. For this purpose, we must take into account three aspects: the extension of a given plural marker in the paradigms and among the verbs (i.e. its productivity), its morphonological transparency (i.e. if it is identifiable with another morpheme, especially from the NP declension), and, if recognizable, the productivity/transparency of the homonymous morpheme itself, as a clue to its ancientness. Likewise, we should bear in mind that the marking of the plurality probably emerged first in the 3rd person forms, and

28 Although this scheme only completely fits with the standard variety, it can hold roughly valid for almost all of the dialects.

then extended to the others (only if S or P), although these already made the distinction lexically (*n-*, *h-* [SG] vs. *g-*, *z-* [PL]).

If we compare the extension of each pluralizer (*-it-, *-za, *-de*), we soon observe a big difference between them. The first (*-it-*) is only used with the transitive auxiliaries **edun* (also synthetic), **ezan* and **iron*. Its use with the verb *-i-* (as a suppletive one for **edun* in the allocutive conjugation, or —more dialectally restricted— as a ditransitive auxiliary) can be understood to be a later extension of the form. As we have seen, its origin remains unexplained, but we can guess how ancient it is by considering its shape (VC), and the fact that it is always prefixed.

With respect to the other two, they are more widely employed, and they are found in a rather sharp distribution. On the one hand, **-de* has the monopoly of A and R plurality marking (in all dialects), and is also used as an S pluralizer of the auxiliary **edin* and —in most dialects— of the verb *egon*. In addition, it is the sole (re)pluralizer of the new (and pleonastic) 2nd p. pl. Hence, it can pluralize all of the Ss and also all of the Ps for that person. On the other hand, **-za* is the most widespread pluralizer for patients (in fact, for all the verbs except those with the exceptional *-it-*), and for subjects.

A final question concerning these S/P pluralizers is the syntactic factor which triggers their appearance. In the easternmost dialects and old western texts, there seem to be two conditions under which they can appear: countable plurality (two or more) and definiteness of the corresponding NP (thus, taking an S as an example: *bi gizon etorri da* 'two men have come' vs. *bi gizon-ak etorri dira* 'the two men have come'). This seems to be the old state of affairs (cf. *HB*: 221). As Martínez-Areta reminds me, since the plurality of the 1st and the 2nd plural pronouns also inherently contains these two features, we could suppose that the 3rd person S/P plural marking began when the 1st and 2nd p. pl. pronouns were indexed on the finite verb (as prefixes). This brought about a need to analogically create a 3rd p. pl. agreement marker expressing the same plurality, which in this case was created by suffixing. Later, these markers spread to finite forms with 1st and 2nd p. pl. pronoun agreements for S/P, where they were actually redundant. In a dialectal period, they also penetrated into contexts which did not fulfill the two aforementioned conditions, such as indefinite ones (hence, → *bi gizon etorri dira* 'two men have come').

9.4. Tense-Aspect-Mood

In this section, I shall discuss the tense and mood distinctions made by the Basque finite verb in its synthetic conjugation. Aspect will only be mentioned as

subdued to tense, since synchronically it is no longer a relevant category in synthetic conjugation, that is, there is no morphology for any aspectual contrast,[29] and diachronically it was more related to the semantics of each verb than to their morphological shape. Similarly, we should note that the system presented here is not found as a whole in any modern dialect, but was definitely the old one, which can be traced back to the oldest attestations. Hence, we will indicate the main diachronic differences where necessary.

9.4.1. Tense and Aspect

As has been seen throughout this work, the synthetic verb can express three basic tenses: present, past, and the so-called *hypothetical* tense.[30] They are characterized by two features: 1) the use of specific morphemes, and 2) a particular morpheme order. A remarkable fact is the formal connection between the past and the hypothetical tense.

9.4.1.1. Specific morphemes

9.4.1.1.1. Present

The present tense is formed by the prefix *da-*, or its allomorphs *d-* (only with 3rd p.), *-a-* and *-ø-* (rarely and only with 1st-2nd p.), added to the verb root, thus forming a stem to which person agreement markers are prefixed or suffixed:

(9)	a. *ø-**da**-tor*	b. *ø-**da**-kar-t*	c. *ø-**d**-oa* (< ***da**-oa*)	d. *ø-**d**-u-t* (< ***da**-du-t*)
	3-PRS-come	3-PRS-bring-1SG	3-PRS-go	3-PRS-have-1SG
	‘(S)he comes’	‘I bring it’	‘(S)he goes’	‘I have it’

(10)	a. *n-**a**-tor*	b. *n-**a**-kar-ø*	c. *n-**ø**-oa* (< *n-**a**-oa*)	d. *n-**a**-(*d)u-ø*
	1SG-PRS-come	1SG-PRS-bring-3	1SG-PRS-go	1SG-PRS-have-3
	‘I come’	‘(S)he brings me’	‘I go’	‘(S)he has me’

29 In periphrastic constructions, the lexical verb does carry the aspectual meaning, and it distinguishes between imperfective (*-t(z)en*), perfective (*-tu*) and prospective (*-ko*) (see §8.3).

30 This tense cannot be used in its bare form (i.e., without complementizers or modal markers) but serves as the base for the hypothetical counterpart of the other tenses (see Tables 9.2.1, 9.2.2 and 9.2.3).

Indeed, the fact that *d-* always (and only) appears with the 3rd person had led grammarians to interpret this consonant as an agreement marker (cf. Lafon 1980 [1944], among others), but, since Trask (1977), we consider all these variants to be allomorphs of a single morpheme: the present tense/imperfective aspect *da-*. In fact, Trask (1977) also took a step further in the diachronic explanation of the 1st and 2nd p. markers, proposing that *na-*, *ha-*, *ga-*, *za-* come from */ni-da-.../, */hi-da-…/, */gu-da-.../, */zu-da-…/; that is, clitic pronouns fused to the tense prefix of the "impersonal" verb.

Regarding the origin of the present marker *da-*, de Rijk (1995a) proposed reconstructing it as **dan*, which would coincide with the *-dan-* element that appears in some temporal adverbs such as *orain-dan-ik* 'from now on' [now-*dan*-ABL].[31] De Rijk believed that this was the old Basque word for 'now'.[32] However, in recent papers, Lakarra (2006b, 2008a) has recovered and defended Trask's idea (1977). According to this, the morpheme *da-* has originally had a progressive-imperfective value (thus, not so easily relatable to de Rijk's reconstruction), which later changed to its modern present sense. This hypothesis is supported by Lakarra's own reconstruction, according to which a posture verb **dar* 'sit' grammaticalized as an aspect marker. This **dar* 'sit' would be "present" historically as the root of the non-finite verb *jarri* (< **e-darr-i*) [sit.PTCP]. Such an evolution has plenty of parallels in grammaticalization theory (Heine & Kuteva 2002).

9.4.1.1.2. Past

The past tense has a more complex structure, because of the above mentioned variation in the collocation of the agreement markers. The basic stem has a prefix *e-* (or > *ø-*) when the S/P agreement is 3rd p. (that is, zero) (11a-b, 12c), or *en-* (> *in-*)[33] when the S/P is other than 3rd person (11c, 12a-b). To this stem, the suffix *-(e)n* (or the dialectal variant *-an*) is added. The prefix seems to be a kind of non-present marker, since it also appears in the hypothetical tense; thus, the suffix is the actual bearer of the past sense. The agreement prefixes are the same as those of the present tense, but when the word-initial agreement marker is 3rd p. (and recall that this marker is *ø-*), a *z-* is added in all dialects except in Biscayan and Alavese: *z-e-go-en* (B. *egoan*), *n-en-go-en* (B. *nengoan*). However, in those dialects which add the element *z-*, this has been traditionally interpreted as

31 But cf. the analysis given above: *-da-n-ik* [*da*-INES-ABL].

32 The historical and modern word for 'now', *orain*, is supposed to contain a Latin loanword, from *hora* 'time, hour' (*FHV*: 141).

33 Or *-ind(e/a)-*, a variant resulting from analogical leveling (Oregi 1974).

the agreement of the 3rd person (as has occurred with the present prefix *d(a)-* and the hypothetical *l(e)-*; despite this, I have glossed them with their original meaning). Other agreement suffixes are inserted before the final *-(e)n*. Here are some examples:

(11)	a. ***z-ø-e-torr-en***	b. *n-**e**-karr-ø-**en***	c. *n-**in**-du-ø-**en*** (< **n-en-du-en*)
	PST-3-NPRS-come-PST	1SG-NPRS-bring-3-PAST	1SG-NPRS-have-3-PST
	'(S)he came'	'I brought it'	'(S)he had me'
(12)	a. *n-**en**-torr-**en***	b. *n-**inde/en***[34]*-karr-zu-**n***	c. *n-**ø**-u-ø-**en*** (< **n-e-du-en*)
	1SG-NPRS-come-PST	1SG-NPRS-bring-2PL-PST	1SG-NPRS-have-3-PST
	'I came'	'You brought me'	'I had it'

9.4.1.1.3. Hypothetical

Finally, we should discuss the so-called hypothetical tense, which is not actually an autonomous tense. Moreover, it only overtly appears when the word-initial morpheme corresponds to the 3rd person, by means of the *l-* prefix. In fact, it has the same stem as the past tense (*-e(n)-CVC-*), with the difference being that it does not bear the past morpheme *-(e)n*, and it must carry some modal morpheme: the conditional prefix *ba-* 'if' (protasis), the potential morpheme *-ke* (apodosis, possibility), or the subjunctive *-(e)n* (hypothetical wish). There is only one exception, which is archaic and no longer available in the language: the "*l-* past indicative". This form was used to express the past tense in subordinated clauses, as in Laz. *Errespuesta emaiten zidan **lizala*** [instead of *zala*] *neska onestua* 'She (always) answered me **that she was** an honest girl'.

As can be observed in the following examples, this form always bears either the suffix *-ke* or the prefix *ba-*:

(13)	a. ***ba**-n-**en**-tor*	b. ***ba**-n-**e**-kar-ø*	c. ***ba**-n-**in**-du-ø* (< *banendu*)
	if-1SG-NPRS-come	if-1SG-NPRS-bring-3	if-1SG-NPRS-have-3
	'If I came'	'If I brought it'	'If (s)he had me'

34 The form *-inde-* belongs to the eastern and the standard varieties, whereas *-en-* (the original one) is used in the central-western part of the country. The first one resulted from an analogical leveling with the paradigm of **edun*, whereby the *-d-* of such forms was reinterpreted as part of the prefix instead of the root (*n-en-du-en* > *n-ind-u-en*), also due to an analogy with the present tense forms, where the root is *-u-*, not *-du-* anymore (**na-du* > *na-u* '(s)he has me', **da-du* > *d-u* '(s)he has it', etc.).

(14)	a. ***l-ø-e-tor-ke***	b. *n-**inde/en**-karr-**ke**-zu*	c. *n-**ø**-u-**ke**-ø* (< *neuke*)
	HYP-3-NPRS-come-IRR	1SG-NPRS-bring-IRR-2PL	1SG-NPRS-have-IRR-3
	'(S)he would come'	'You would bring me'	'I would have it'

9.4.1.2. Morpheme order and ergative displacement

As it has been pointed out, there is a quite evident similarity between the two non-present tenses. The similarity lies not only in the TAM morpheme *-e(n)-*, but also in the order of agreement markers. In this respect, both of them show a split pattern in transitive verbs: if P is 1st/2nd p., then it follows the same model as the present tense, while if P is 3rd p., the pattern is reversed. Since the only overt agreement marker in such forms is that of the A (1st/2nd person), this phenomenon is synchronically named *ergative displacement* (Laka 1988, 1993), and is illustrated in Figure 9.2.

Present pattern	S/**P**-V-**A** (***n**-akar-**ø***, ***ø**-dakar-**t***)
Non-present pattern I	S/**P**(1st&2nd)-V-**A** (*-**n**-indekar-**ø***-)
Non-present pattern II	S/**A**-V-**P**(3rd) (*-**n**-ekar-**ø***-)

Figure 9.2. Position of A and P markers according to the ergative displacement phenomenon

In synchronic morphosyntaxis, some attempts have been made to explain this alternation (Laka 1988, 1993, Ortiz de Urbina 1989). As for its diachronic origin and development, one path of research is that of Heath (1977), continued by Aldai (2000). An opposing viewpoint is taken by Martínez-Areta (2012) (see the final Section §9.8).

9.4.1.3. Inherent Aspect: past, present and future

This subsection deals with the crucial role that the inherent aspect of verbs used to have regarding these tenses (it no longer does nowadays) in the semantics of the Basque synthetic conjugation.

As already mentioned, the modern synthetic verbs do not differentiate at all between perfective and imperfective aspects. More precisely, they always express the imperfective or progressive one. However, this is not the case in the oldest texts of different dialects. At that time (ca. 16th c.), the particular meaning of each verb conveyed a specific aspectual behavior (Lafon 1980 [1944]).

Thus, if a verb expressed an action which has an inherent endpoint, i.e. one which we expect to finish at some point, its synthetic past forms used to carry a perfective meaning (15); and *vice versa*, if the action could be seen as not hav-

ing a built-in endpoint, i.e., if it was continuous and did not end at a particular point, the synthetic past denoted imperfective aspect (16):

(15) a. *nentorren* 'I came' (→ 'I was coming')
b. *nekarren* 'I brought it' (→ 'I was bringing')

(16) a. *nengoen* 'I was (staying)' (→ 'id.')
b. *nuen* 'I had it' (→ 'id.')

To express imperfectivity, verbs like (15) had to be conjugated periphrastically, or, in some cases, the morpheme *-ke* also served to change the aspectual nuance (as in *nekarkeen*, cf. (15b)).

With respect to present forms, they had an even greater variability in meaning, as is mostly attested in Old Biscayan. For example, some verbs (such as *egin* 'to do') could express the future with the present bare form (17a), or by adding the subjunctive *-(a/e)n* (17b) or the *-ke/-te* suffix (17c):

(17) a. *daik* 'thou will do' (RS), *eztegik* 'it will not tear for you' (Oih.)
b. *ax**a(n)*** 'thou will be' (RS), *dagida**n*** 'I will do' (Gar.)
c. *goa**kez*** 'we will go' (RS), *axa**te*** 'thou will be' (RS)

9.4.2. Mood

Scholars usually describe five moods for the general verb system: indicative, conditional, potential, subjunctive, and imperative (*HB*, *GB*, de Rijk 2008). However, there is not always such a clear morphological distinction or characterization in the synthetic conjugation; this distinguishes morphologically up to five moods, but defectively, depending on the person agreement and the tense. Tables 9.2.1, 9.2.2 and 9.2.3 below summarize the features of each mood, combined with the three tenses, depending on the S/A agreement and exemplified by the verbs *etorri* 'to come' and *ekarri* 'to bring'.

In them, we can see some differences regarding the imperative(-like) mood(s), which can be summarized as follows. On the one hand, the 3rd person has a distinctive form for the jussive for virtually all the synthetic verbs, along with the complementizer suffix (*-ela*) which can be used with the same meaning (cf. Goenaga 1996: 127). The 2nd person imperative forms of the transitive verbs are also unique (e.g. *ekark*, although forms like this have usually been interpreted as pseudo-synthetic or contracted forms, cf. Oyharçabal 2003: 279, and Mujika 2011).

Table 9.2.1. Mood marking structure of synthetic verbs with 3rd p. S/A

	Present (*da-*)	Past (*ze-/-en*)	Hypothetical (*le-*)
Indicative	*dator* *dakar*	*zetorren* *zekarren*	—
Conditional protasis (*ba-*)	*badator* *badakar*	*bazetorren* *bazekarren*	*baletor* *balekar*
Cond. apodosis / potential (*-ke*)	*datorke* *dakarke*	*zetorkeen* *zekarkeen*	*letorke* *lekarke*
Subjunctive (*-en*)	*datorren* *dakarren*	*zetorren* *zekarren*	*letorren* *lekarren*
Jussive (*be-*, *-ela*)	*betor*, *datorrela* *bekar*, *dakarrela*	—	

Table 9.2.2. Mood marking structure with 1st p. S/A

	Present	Past	Hypothetical
Indicative	*nator* *dakart*	*nentorren* *nekarren*	—
Conditional protasis	*banator* *badakart*	*banentorren* *banenkarren*	*banentor* *banenkar*
Cond. apodosis / potential	*natorke* *dakarket*	*nentorkeen* *nekarkeen*	*nentorke* *nekarke*
Subjunctive / (co)hortative	*natorren* *dakardan*	*nentorren* *nekarren*	

By contrast, the intransitive imperatives of the 2nd person and all the (co)hortatives of the 1st person have been transferred from another mood. Thus, the 2nd person form *hator* 'come!' is the same as that of the present indicative *hator* 'thou art coming', and the 1st person forms like *natorren* 'let me come' and *dakardan* 'let me bring it' are actually subjunctive forms.

Nevertheless, it must be taken into account that nowadays the synthetic subjunctive is no longer in use, except in a few fossilized expressions (such as the cohortative *goazen* 'let's go!') and in formal language (*demagun...* 'let us assume...', *noan harira* 'let me get to the point', lit. 'let me go to the thread'). The present forms with *-ke* (or its allomorph *-te(ke)*) have also disappeared from the western speech (except those of the irrealis auxiliaries **edin* and **ezan/*egin*) and they have reduced their range of use in the eastern varieties, since now they are only used to express uncertainty.

Table 9.2.3. Mood marking structure with 2nd p. (M) S/A

	Present	Past	Hypothetical
Indicative	*hator* *dakark*	*hentorren* *hekarren*	—
Conditional protasis	*bahator* *badakark*	*bahentorren* *bahekarren*	*bahentor* *bahekar*
Conditional / potential	*hatorke* *dakarkek*	*hentorkeen* *hekarkeen*	*hentorke* *hekarke*
Subjunctive	*hatorren* *dakarran*	*hentorren* *hekarren*	
Imperative	*hator* *ekark*	—	

By contrast, the intransitive imperatives of the 2nd person and all the (co)hortatives of the 1st person have been transferred from another mood. Thus, the 2nd person form *hator* 'come!' is the same as that of the present indicative *hator* 'thou art coming', and the 1st person forms like *natorren* 'let me come' and *dakardan* 'let me bring it' are actually subjunctive forms.

Nevertheless, it must be taken into account that nowadays the synthetic subjunctive/(co)hortative is no longer in use, except in a few fossilized expressions (such as the cohortative *goazen* 'let's go!') and in formal language (*demagun...* 'let us assume...', *noan harira* 'let me get to the point', lit. 'let me go to the thread'). The present forms with *-ke* (or its allomorph *-te(ke)*) have also disappeared from the western speech (except those of the irrealis auxiliaries **edin* and **ezan/*egin*) and they have reduced their range of use in the eastern varieties, since now they are only used to express uncertainty.

It is also remarkable that nowadays the most frequently employed verbs (*izan* 'to be', **edun* 'to have') lack synthetic imperative forms, and hence there is no longer any form like *haiz* 'be!', *euzu* 'have it (you)!' or *b(i)u* 'let him/her have it!', and it is only in the highest register that *biz* 'let it be!' can be used.

Moreover, some analogical leveling has occurred, at least in the dialectal written attestations, in the paradigm of the imperative-like forms, that is, in those forms which had not originally had a genuine imperative form, e.g. *nazala* [3sg.A-1sg.P] (with the complementizer *-(e)la*). In some eastern varieties, this auxiliary changed to *benaza* according to 3sg.-patient forms (*beza*, with *be-*), and, in some western ones, it became reshaped as *naza*, following the pattern of 2sg.-agent forms (*nazazu*, without *-(e)la*).

Indeed, *bare-present-shaped* transitive imperatives[35] such as *nazazu*, finite form of the auxiliary ***ezan* (also west. *nagizu*, from *egin*), are also an exception today, because there is no other verb which has them. However, we can assume that at some point such forms were common, as evidenced not only by the quoted forms, but also by older ones such as *nafinzu* 'put me!', *nanzuzu* 'hear me!', *nakusu* 'see me!' or *(zeure besoan) narazazu* 'let me lie (on your arms)!' from Lazarraga's manuscript, or *(ukenduru) negizu* 'make me (owner)!' from a poem by Oihenart.

To sum up, we can say that there are four morphemes which mark mood: *ba-* (conditional protasis), *-ke* (potential, conditional apodosis), *-en* / *-ela* (subjunctive, along with imperative-like uses) and *be-* (jussive, only with the 3rd person). Moreover, there is a *l(e)-* prefix which nowadays serves (when the subject or the agent is the 3rd person) to mark non-indicativeness. However, as I showed above, in old texts we find an archaic use, whereby, in subordinated clauses corresponding to main clauses with a finite verb in a non-present tense, the subordinated finite verb begins with that *l-*, e.g. Dch. *[Gaitzez **lago**la] enzun nuien* 'I heard [that **he was** angry]' (instead of *zegoela*).

9.5. Valency and diathesis

Another aspect that we have to consider when analyzing the verbal morphology of a language is the way in which the different types and alternations of the valency are reflected in it. In fact, it is likely that they have no influence (or nearly none) on the morphological structure of the verb. If we take the Spanish verb (and the same would hold true for most —if not all— Indo-European languages), we can see that both transitive and intransitive verbs only agree with the corresponding subject, and that there is no distinctive marker for each type of valency (*corr-o* 'I run', *com-o* 'I eat', cf. their respective infinitives *corr-er* 'to run', *com-er* 'to eat'). However, it is well known that Basque expresses, morphologically and in a single verbal form, all the arguments of the clause —whose number and relative order tell us the valency of verb (cf. Figure 9.1 and Table 9.1)—, including the presence of an indirect object, if there is one (through the *dative flag* or the alternation of the *tense vowel*).[36] Lafon (1980 [1944]: 400-402) emphasized this richness of agreement markers, but was unconvinced by Gavel and Lacombe's explanation according to which such rich-

35 Cf. the intransitives *hago* 'you are staying' or 'stay!, wait!' and *zoaz* 'you are going' or 'go away!'.

36 In SB, intransitivity can also be marked by the concrete allomorph of the non-present marker, although it does not seem to have had that function originally.

ness was attributed to the influence of Gascon and Spanish, due to the clitic datives of those languages.

The eminent French linguist Lafon (1980 [1944]: 18-24) distinguishes between two types of verbs according to their transitivity, specified by the presence/absence of an agent (*Nork*). In the first group (intransitives), we find the *Nor* [ABS] and *Nor-Nori* [ABS-DAT] verb types, in the second one (transitives) we find the *Nor-Nork* [ABS-ERG] and *Nor-Nori-Nork* [ABS-DAT-ERG] verb types.[37] Some verbs are only conjugated as Nor (but the majority of them are also conjugated as Nor-Nori) and some are only conjugated as Nor-Nork; some others can be conjugated as Nor-Nork and as Nor-Nori-Nork; and a few of them are only of the Nor-Nori or Nor-Nori-Nork type. Another possible classification can be made according to the number of agreements which the verb accepts. Thus, we would have 8 types: of one person, of two persons (Nor-Nori and Nor-Nork), of three persons, and the allocutive forms (§9.6) which belong to each of the previous types (with two, three, and up to four agreements, respectively).

On the other hand, and as Lafon (1980 [1944]: 24-25) warns, there is no verbal root (leaving aside two exceptions)[38] which can be synthetically conjugated in both an intransitive and a transitive way; nor is there any morphologically productive process which can change the (in)transitivity of a verb in its synthetic conjugation. Nonetheless, in some verbs we can see an originally causativizing *-ra-* prefix, whose derivations (which nowadays do not so clearly maintain that semantic connection) differ from the verbs from which they were derived because of their transitivity (cf. *ibili* 'to walk, to move' [intransitive] : *erabili* 'to use, to handle' [transitive]).

The non-finite forms of the periphrastic constructions, however, can take the *-arazi* suffix to form a causative, and certain verbs can be conjugated both with intransitive auxiliaries and with transitive ones to change their valency (cf. *hil da* '(s)he has died' vs. *hil du* '(s)he has killed (someone)').

Thus, the Basque verb distinguishes four valency types: monovalent intransitive (§9.5.1), bivalent intransitive (§9.5.2), transitive (§9.5.3) and ditransitive (§9.5.4). In the transitive type, however, there are some "monovalent transitive verbs", which are morphologically transitive verbs but with only one argument at clause level (they look like simple transitives). An example of such verbs is *urak diraki* 'the water is boiling', in which the sole argument of the verb 'boil'

37 *Nor* means 'who (abs.)' in Basque, *nork* 'who (erg.)', and *nori* 'to whom'. In Basque grammar, these terms are used to label each verb type according to the arguments they agree with.

38 Cf. *natxiko* : *datxikat*, both 'I adhere to it/him/her', and *hautso* : *deutsak/n* , both 'thou hold on to it/him/her'.

takes the ergative case *-k*, just like the agents of the common transitive predicates; cf. *ura doa* 'the water is going', with absolutive *-ø*.

To conclude this section, I should make it clear that even if Basque can form passive-like constructions, these are often considered to be unnatural by Basque grammarians, arising from the influence of French and Spanish. Instead of making voice distinctions, Basque resorts to alternations between the four verb types (cf. §9.5.5). However, based on the morphology of the Basque verb, some authors have argued for an alleged *passivity*, whether in historical Basque (Schuchardt 1972 [1893], 1947 [1923], Lafon 1980 [1944]), or at a previous stage of the language (Trask 1977), or they have perceived, in certain features, the absence of diathesis as a grammaticalized category at some point in the past (Martínez-Areta 2012) (see §9.8.2).

In the following subsections, I shall explore each type of valency in more detail, and show how we can derive one from the other. I will also look at some hypotheses for the origin and development of the Basque diathesis system and their implications for the proto-language.

In each subsection, I have included a table to illustrate the different patterns of verbal valency, showing a part of the paradigm of a handful of verbs. In order to simplify the visual appearance of the tables, only the person agreement is highlighted, and not each morpheme.

9.5.1. Nor

Monovalent intransitive verbs, also called *Nor* [who.ABS] verbs, are those which agree with one argument (the subject) that bears the absolutive case marker *-ø*. These verbs express a state, a change of state, or a movement. Two of them are also auxiliaries: *izan* and **edin*, with the latter verb nowadays functioning only as an auxiliary, but with well-attested independent uses.

In addition, there are certain verbs (e.g. *ibili* 'to walk, to move', cf. Spanish *andar*, and *egon* 'to stay, to be at', cf. Spanish *estar*) which have some properties of auxiliaries, like semantic bleaching in sentences such as *abestiak egi**ten dabil*** '**(s)he is** mak**ing** songs' and *irakur**tzen dago*** '**(s)he is** read**ing**'. Another one has been employed (but not anymore) to express the habitual aspect of intransitive clauses: *joan* 'to go', as in *galdu **doa*** '**(s)he usually** get**s** lost'.

The only argument (the subject) of Nor verbs is invariably encoded by a prefix in all tenses and moods. Table 9.3 is an exhaustive list of all the Nor verbs which have at least one attested synthetic form.

Table 9.3. Basque verbs which show at least one Nor synthetic form

**edin* 'to become'	*izan* 'to be'
egon 'to stay, to be at'	*jaiki* 'to stand up, rise (up)'
etorri 'to come'	*jaitsi* 'to go/come down'
etzan 'to lie (down), rest'	*jarri* 'to sit (down)'
ibili 'to walk, to move'	*jaugin* 'to come'
igaran 'to pass', 'to go by'	*joan* 'to go'
irten 'to go/come out'	

From this limited group, approximately half of them are used synthetically in the spoken language (*egon*, *etorri/jaugin*, *ibili*, *izan*, *joan*), depending on the dialect or the verb (for example, from *jaugin* we know of only a few imperative forms for the 2nd persons and another one for the relative of the 3rd person). To this small group, I should add **edin*, which is no longer an autonomous verb, but an auxiliary for *irrealis* moods.

Table 9.4 shows how the intransitive monovalent verb agreement is reflected, distinguishing between two main stems (even if this does not have any influence on the intransitive verbs), and for this purpose I shall take the roots of *izan*[39] 'to be' and *ibili* 'to walk, to move'.

Table 9.4. S marking in intransitive monovalent verbs

Nor	**Stem I (present)**		**Stem II (non-present)**	
1sg	**n**aiz(a-)	**n**abil	-**n**intz(a-)	-**n**enbil-
3sg	**ø**da	**ø**dabil	-**ø**(i)z(a-)	-**ø**ebil-
2pl	**z**ara	**z**abiltza	-**z**ina-	-**z**enbiltza-

As will be shown later, sometimes it is the semantics of the verb which shows its valency and not the morphology, since this can be ambiguous in the 3rd person forms (compare *dabil* '(s)he walks, it works' vs. *dakar* '(s)he brings it').

39 As it is well known, the verb *izan* 'to be' is the most irregular one, not only because of the differences between number and persons, but also within the latter, as is the case with the 3rd person, depending on the tense/mood; however, we can reconstruct the more or less regular 3rd person proto-paradigm as follows: *zen* 'it was' < ****(i?)za**-en*, *biz* 'let it be!' < **be-**iza***, *balitz* 'if it were' < **ba-le-**iza***, *litzateke* 'it would be' < **le-**iza**-teke*. The affricate sibilants of the last two forms are diachronically secondary.

9.5.2. Nor-Nori

The synthetic verbs which have an exclusively bivalent intransitive conjugation (or *Nor-Nori* [who.ABS-who.DAT]) are few and quite defective, since frequently the form of the third person of the indicative is the only one used. Some of them (*atxiki*, *eutsi*, *jarraiki*...) were intransitive, but changed to the transitive type due to reinterpretations owing to their aforementioned morphological ambiguity. For instance, along with the old form *zarraizkit* 'you follow me' with a Nor-Nori morphology, we can also find forms like the ditransitive *darraidazu* 'id.',[40] while on the other hand the archaic imperative *hautso* 'hold it!' has given rise to *eutsiok* 'id.' (cf. also *dautso* 'it holds it' vs. B. ditrans. aux. *deutso* 'it has it to it'; §9.5.4). Table 9.5 is a list of Nor-Nori verbs which take synthetic conjugation in at least one attestation.

Table 9.5. Basque verbs which show at least one Nor-Nori synthetic form

atxiki 'to hold on, to join'	*jario* 'to flow, to pour'
ekin 'to undertake, to start'	*jarraiki* 'to follow, to chase'
eutsi 'to hold (on)'	*jarrugi* 'to find'
izeki 'to catch (fire)'	

In the synthetic forms of these verbs, the subject always appears as prefixed, as in the Nor forms, whereas the recipient Nori is always suffixed, and they can never coincide in the person category (***narrait* 'I follow myself'); that is, they cannot be reflexive (Lafon 1980 [1944]). This is illustrated in Table 9.6.

Table 9.6. S and R marking in intransitive bivalent verbs

Nor	**Nori**	**Stem I (present)**		**Stem II (non-present)**	
1sg	3sg	**n**aki**o**	**n**arrai**o**	**-n**enki**o-**	**-n**inderrai**o-**
3sg	3sg	**ø**daki**o**	**ø**darrai**o**	**-ø**eki**o-**	**-ø**errai**o-**
2pl	3sg	**z**akizki**o**	**z**arraizki**o**	**-z**enkizki**o-**	**-z**inderraizki**o-**

In Table 9.6, we can see a common feature for almost all of the verbs of this type: they have a morpheme which we call the *dative flag* (*-ki-*, *-(g)i-*, *-ts-*). In fact, this morpheme (mostly *-ki-* in the standard variety) can be applied to many

40 Obviously, when they change to the Nor-Nori-Nork type, it only affects verb agreement and the case marker taken by the corresponding NP, not the number of arguments in the clause (*zu-**ø** niri zarraizkit* → *zu**k** niri darraidazu**u***). Thus, the absolutive one is left undetermined, and *unmarked*, for it is always the 3rd person (cf. §9.5.4).

verbs from Table 9.3 to add dative agreement. Examples include *egon* 'to stay' and *etorri* 'to come': *dago* 'it is at' → *dagokio* 'it corresponds to it', *dator* 'it is coming' → *datorkit* 'it is coming to me', 'it is [good/bad] for me'.

The so-called *dative flag* precedes the agreement morpheme of the recipient almost without exception, and its function is to increase the valency of a verb that does not have the dative in its simplest argument structure.

Schuchardt (1972 [1893]) considered the oldest form to be *-ki*, which he related to the sociative *-kin* and dative *-i* suffixes (ibid.: 29). From this form, the other forms were derived: *-i* as a result of the loss of *-k-* (ibid.: 46), and *-ts-* as a result of "mutual assimilation" (ibid.: 44). However, these changes would represent a mismatch with the current standard historical phonology of the Basque language.

According to Trask (1995b: 231), and leaving aside the *-ts* morpheme, the allomorphs *-gi*, *-i*, *-ki* seem to have been derived from **-gi* in the same way as the pluralizers *-de*, *-e*, *-te* come from **-de* (cf. §9.5.4.1.1). The variant with the voiceless consonant seems to have been interpreted (and generalized) in certain contexts as a part of the plural marker (**-z-gi-* > *-z-ki-* → *-zki-*, §9.3.2). There is a non-standard form, historically common to all dialects, which is the *-k-* alone, and which Schuchardt (1972 [1893: 46]) identified as a variant of *-ki* before the dative *-o* for the third person in forms such as *dabilko* (< *dabilkio*), *dagoko* (< *dagokio*).

As Schuchardt noted (1972 [1893: 50]), there are examples from many dialects and early periods in which the *dative flag* is not used. Instead, in some cases the consonant of the agreement marker that would follow it appears to be "strengthened", that is: *doa**k**i**z**u/**g**u* ~ *doa**t**z**u*/*k**u*** 'it is going to you/us', *doa**k**i**d**an* ~ *doa**t**an*** 'which is going (for me/my...)'. Lafon (1999 [1961]: 426) suggested an analysis of *-k-* and *-i-* as two originally different morphemes, and proposed that the initial consonants of the agreement markers were "strengthened" after that *-k-* (thus *doaku* < **doa-k-gu*). In view of the cited examples, it is more likely that the *-i* of the *dative flag* (*-ki*) was simply lost when adding the agreement marker, especially in the longest verbal forms, following the old compounding or derivational rule (cf. *FHV*: 125-128) according to which the *-i* (and the *-u*) of the first element of a compound or nominal derivation (or even of a [noun + adjective] phrase) was lost if that element had more than one syllable. That apocope could have occurred more easily, also with other vowels, when the first element had three or more syllables (cf. *harpe* 'cavern', from *harr**i*** 'stone' + *pe* 'under', *itsasgizon* 'seaman', from *itsas**o*** 'sea' + *gizon* 'man'). That is, according to this rule, from **doaki* plus the agreement markers, the forms *doak(i)-o*, *doak(i)-gu*, *doak(i)-zu* would emerge, and following Basque phonological rules we would obtain *doako*, *doaku* and *doatzu*.

As far as the *-ts* marker is concerned, it appears more frequently in finite forms of the western dialect (cf. ditransitive auxiliary *eutsi*, or three-person forms of the verb *eduki* 'to keep, to have' like *daukotsot*, presumably created by analogy), but its general extension is well documented both through synthetic forms of the verb *-io-* 'to tell' (*diot* 'I tell sth.' vs. *diotsot* 'I tell him/her sth.') and through non-finite forms of certain verbs, such as *eraku**ts**i* 'to show' < **e-ra-kus-ts-i* (old causative of *ikusi* 'to see') or *iraka**ts**i* 'to teach' < **i-ra-kas-ts-i* (old causative of *ikasi* 'to learn') (Schuchardt 1972 [1893: 46], Trask 1995b: 229).

9.5.3. Nor-Nork

In this subsection, we will explore the field of transitive valency, the second of the groups mentioned at the beginning of this section. The mono-transitive type (patient-agent) is historically more numerous among the verbs with some trace of the synthetic conjugation, although in the spoken language of today only a handful of them are used. Table 9.7 is a list, which if not exhaustive is close to it, of the Nor-Nork verbs with at least some attested synthetic forms.

Table 9.7. Basque verbs which show at least one Nor-Nork synthetic form

edasi 'to talk'	*ezarri* 'to put'
edeki 'to take off'	*idoki* 'to take away, out'
eduki 'to hold, to have'	*igorri* 'to send'
**edun* 'to have', 'trans. aux.'	*ihardetsi* 'to reply'
egin 'to make'	*iharduki* 'to fight', 'to argue'
egotzi 'to throw'	*ikusi* 'to see'
ekarri 'to bring'	*-io-* 'to tell'
entzun 'to listen'	*ipini*, *ifini*, *imini* 'to put'
erabili 'to use'	*iraitzi* 'to know', 'to reject'
eraitsi 'to take/pull down'	*irakin* 'to boil'
eraman/eroan 'to carry'	*iraun* 'to last'
eratzan 'to lay down'	**iron* 'trans. irr. aux.'
erauntsi 'to pour out/down'	*irudi* 'to seem'
erosi 'to buy'	*jakin* 'to know'
erran/esan 'to say'	*jo* 'to hit'
ezagun 'to know'	*sinetsi* 'to believe'
**ezan*'trans. irr. aux.' (< 'to obtain (?)')	*utzi* 'to leave'

By and large, the P agreement morpheme is prefixed and the A is encoded by a suffix. However, we must bear in mind what I mentioned in §9.3 and §9.4 about the "TAM split" shown by these verbs when the P belongs to the 3rd person.

With the present stem, they behave as I have just explained, but with the non-present stem the values change: the prefix is used for the A and the suffix for the P. As we can see in Table 9.8.2, the *n-* prefix can refer to both the absolutive argument (*-nindekarzu-*) and to the ergative one (*-nekar-*), depending on whether the absolutive agreement marker belongs to the 3rd person or not.

Table 9.8.1. P and A marking in transitive bivalent verbs of the stem I type

Nor	**Nork**	**Stem I (present)**	
1sg	2pl	**n**akar**zu**	**n**arabil**zu**
3sg	1sg	**ø**dakar**t**	**ø**darabil**t**
2pl	3sg	**z**akartza**ø**	**z**arabiltza**ø**

Table 9.8.2. P and A marking in transitive bivalent verbs of the stem II type

Nor	**Nork**	**Stem II (non-present)**	
1sg	2pl	**-n**indekar**zu-**	**-n**inderabil**zu-**
1sg [nork]	*3sg [nor]*	**-n**ekar**ø-**	**-n**erabil**ø-**
2pl	3sg	**-z**indekartza**ø-**	**-z**inderabiltza**ø-**

Some of the verbs in the list above contain the old causative prefix *-ra-*, which increases the transitivity of the original root. Thus, from *joan* 'to go' (< **e-oan*), depending on the dialect, we have *eraman* or *eroan* 'to carry' (both from < **e-ra-oan*); from *etzan* 'to lie' → *eratzan* 'to lay down'; from *ibili* 'to walk' (< *ebili*) → *erabili* 'to use', and so on.

Schuchardt (1972 [1893: 40]) accounted for *iraun* 'to last' (and Lafon 1980 [1944]: 427 accepted this interpretation) as a derivation of this morpheme from the verb *egon* 'to be (at)' → **e-ra-gon* (for the loss of the consonant and the diphthongization of the vowels, cf. *dago* 'it is (at)' → *daude* 'they are (at)'). This explanation presents at least one phonological complication;[41] that is, the

41 We can also raise another (previously mentioned) issue: *iraun* is morphologically transitive but semantically intransitive, so it only accepts one argument (but in the ergative!). Schuchardt explains, or suggests, that the original meaning could be understood as 'make stay, make rest, make continue (being)', and that from reflexive expressions which would become extremely common, such as *nik neure burua/egoera irauten* (< **eragoten*) *dut* 'I make remain, continue myself/my situation', the construction *(nik) badiraut* 'I last' developed through ellipsis; cf. Michelena's identical explanation (1987 [1972d]: 345-348) for the verb *irten* 'to leave, to come out', which in western dialects is conjugated as transitive, and in the rest of the dialects has been adapted to the intransitive conjugation to which it corresponds semantically. According to Michelena, the variants *irten, urten,*

change from *e-* to *i-*, since this change has only occurred in one particular context: when there is another high vowel in the following syllable (cf. *ebili* > *ibili*, *ekusi* > *ikusi*). It is true that some verbs have an *i-* in all the dialects, and hence it is original (*izan*, *ikasi*, *izeki*), but Michelena (*FHV*: 105) interpreted this as being a different morpheme from the *e-*, since it does not disappear when conjugating the verb (*ibili* → *dabil*, *ekarri* → *dakar* vs. *izan* → ***n**ai**z***, *izeki* → ***d**i**z**eka*). Its origin and meaning remain a mystery but, whatever the case, if we accept Schuchardt's etymology for *iraun*, it might be better reconstructed as **(e-?)i-ra-gon*.

Some other verbs recently reconstructed by Lakarra (2008a) are formed by the morpheme**da-*, which would work as a kind of applicative, although, in the case of the synthetic verbs of the Nor-Nork type that etymologically contain it, we cannot tell the difference between the transitivity of the base and the derived one: **e-zun* (> *entzun*) 'to hear' → **e-da-zun* > **ezadun* > *eza.un* > *ezagun* 'to know' (Lakarra 2009a: 330, fn. 87).

9.5.4. Nor-Nori-Nork

There are few three-person verbs or ditransitives in the synthetic conjugation, and at least four of them (all auxiliaries) are historically limited to certain dialects: **eradun* in the east and *eutsi* in the west are synonyms, and are equivalent to *-i-* in its use as an auxiliary, and the High Navarrese **erazan* (Aldai 1999) is equivalent to **ezan* —western *egin*— with the dative. As for the rest, only *iritzi* and *eman* are used nowadays (very defectively, and in the formal register), whereas the others are only attested in older texts. Table 9.9 shows a list of the Nor-Nori-Nork verbs with at least some attested synthetic forms.

Table 9.9. Basque verbs which show at least one Nor-Nori-Nork synthetic form

egin 'to make', 'ditrans. irr. aux.'	*eutsi* 'to hold', 'ditrans. aux.'
eman 'to give'	**e(ra)zan* 'ditrans. irr. aux.'
engu- 'to give'	*-i-* 'ditrans. aux.'
**eradun* 'to give', 'ditrans. aux.'	*i(n)-* 'to give'
eragon 'to get busy with'	*irakatsi* 'to teach'
erakarri 'to attract'	*iritzi* 'to think of', 'to be called'
erakutsi 'to show'	*urgatzi* 'to help'
erau(n)tsi 'to attack, to beat'	

erten come from **eurten*, which in turn would be the imperfective non-finite form of *egorri* 'to send (out)', which, if used in a reflexive way in clauses such as *neure burua egorten dut* 'I send out myself', would give the current western type *irten dut* 'I left'.

Half of these verbs, probably not by chance, contain the prefix *-ra-*, which turns a Nor-Nork verb into a Nor-Nori-Nork one. This issue will be dealt with in Subsection §9.5.4.2 (cf. also §9.5.3). Regarding the rest of the verbs, we should comment on *engu-*, *urgatzi* and *i(n)-* because of their particularity, either due to their very few attested forms (*engu-*, *urgatzi*), or the uses in which they have been grammaticalized (*i(n)-*).

The verb *urgatzi*, which in its periphrastic conjugation is documented in Old Biscayan and in the eastern dialects (especially in Souletin), has only left us three attestations of its synthetic usage, all of which are in a probably fossilized expression. Dechepare writes *Ieinkoak drugatzula* 'may God help you', and in Voltoire's dialogues we find the two variants *Iainkoa trugatzula/trogatzula*. Here, the original *-k* of the ergative has merged with the *d-* of the verb, which would reflect not only the shandi phenomenon that lies behind the *t-*, but also a loss of awareness as to the meaning and the morphological structure of that finite form.

Finally, the verb *i(n)-* —once common to all dialects— which has survived up to the present day in certain expressions, is well attested in many tenses and moods, but its etymology has always been a mystery, both because of its shortness and its alternating forms (*i-* ~ *in-*). There have been many proposals, and although some scholars have got close to a possible solution, Lakarra is the only one who has hit upon a satisfactory answer. Vinson believed that it was a specialization of *egin* used ditransitively, but Schuchardt (1947 [1923]: 46), who did not agree on that point, proposed that in verbs such as *iguzu* 'give it to us!' there was no verbal root, and that its meaning would be something like 'it from you to us', and therefore the *i* would be related to the *dative flag*. The Austrian linguist was not far from the now commonly accepted view, but he did not relate it to the third person auxiliary *-i-*, almost identical to *-i(n)-* in its present indicative forms. However, it seems that Lafon accepted Vinson's idea, and he gave examples from Dechepare, in which, according to him, the *e-* from *egin* was reflected (Lafon 1980 [1944]). Later, de Rijk (1998 [1985]) suggested the possibility that, in fact, the lexical verb behind *-i(n)-* was the verb *jin* 'to come'. Finally, as already mentioned, Lakarra (2008a: 479-482) proposed relating the verb *-i(n)-* 'to give' to the *-i-* of the ditransitive auxiliary, and even to the same vowel of the dative case suffix. According to his hypothesis, the *-i-* auxiliary would simply be a grammaticalization of the most appropriate verb for it, the verb *to give*. On the other hand, the dative marker comes from the same verb root, which would have been grammaticalized as a marker or introducer of re-

cipient arguments in Serial Verb Constructions (SVC):[42] a very common path of grammaticalization in languages with that type of constructions. The etymology of the verb root, taking the *-n-* variants into account, would be **nin* (Lakarra 2008a: 481).[43]

Coming back to the Nor-Nori-Nork type verbs, we must make it clear that nowadays, in these verbs, the P will always be the 3rd person, although there are some auxiliary verbs that have 1st or 2nd person direct objects in Leiçarraga and Azkue. In fact, there is no grammatical factor which prevents these forms from being used, but Bascologists (cf. Gómez & Sainz 1995: 236) have speculated about the possibility that they could have been invented (especially in the case of Leiçarraga, due to translation needs), or created through analogy (in the case of the examples collected by Azkue).[44] This could also be a potential part of the language which has not been productively "used".

In Tables 9.10.1 and 9.10.2, I give some examples of the verbs *iritzi* and *-i-* (ditransitive auxiliary). Note that, since the object is always the 3rd person, the word order change of the Nor and Nork arguments from stem I to stem II will always take place (the dative agreement is always suffixed).

Table 9.10.1. Agreement marking in transitive trivalent verbs of the stem I type

Nor	Nori	Nork	Stem I (present)	
3sg	1sg	2pl	**ø**deriz**ta**z**u**	**ø**di**da**-**zu**
	3sg	1sg	**ø**deritz**o**t	**ø**di**o**-**t**
	2pl	3sg	**ø**derit(z)**zuø**	**ø**di**zu**-**ø**

42 Cf. the explanation that I provide about the origin of the different aspect-tense morphemes (§9.4.1.1 and §9.8.3) and the **-de* pluralizer (§9.3.2.1).

43 Lakarra argues that the verbs which have an *-n* at the end of their root never show it in finite forms (nor do they in some non-finite ones, cf. *e-gi**n*** : *e-gi-te* : *da-gi-t*). Therefore, the *-n* of *indazu*, *indak* —here Lazarraga's *dindala* may be added— could not belong to the end of the root but to the restoration of the nasal from a nasalized vowel in a pre-consonantal position. That vowel was nasalized because an *-n-* between it and the previous vowel (the tense vowel) had been aspirated. Thus, we obtain the **ni-* form which would lack a coda to be fully reconstructed, since we assume (from Lakarra 1995) that Pre-Proto-Basque only had CVC monosyllabic roots. As I have already said, the consonant that does not appear when the verb is conjugated is always *-n*, and that is how we get to **nin*, a root which otherwise had only been proposed to reconstruct the word *ihintz* 'dew', from a partial reduplication plus a suffix (**(n)i-nin-tz*).

44 Cf. what I have said in §9.5.2 about the *dative flag -ts-* of the western dialect.

Table 9.10.2. Agreement marking in transitive trivalent verbs of the stem II type

Nork	**Nori**	***Nor***	**Stem II (non-present)**	
2pl	1sg	*3sg*	-**z**enerizt**(a)ø**-	-**z**enit**/da**ø-
1sg	3sg		-**n**eritz**oø**-	-**n**i**oø**-
3sg	2pl		-**ø**zerit(z)**zu**ø-	-**ø**zi**zu**ø-

9.5.4.1. Dative insertion

We have already seen that an important feature of the Nor-Nori-Nork type verbs is that many of them are derived from Nor-Nork ones, to which a dative agreement has been added (however, there is no Nor-Nori-Nork verb which is derived from bivalent intransitive Nor-Nori verbs by adding an agent). Below, I shall consider some morphological aspects related to this fact (also cf. §9.5.2).

9.5.4.1.1. Dative flag(s)

As is the case with intransitive bivalent verbs, some ditransitive verbs also show certain markers, usually before the agreement of the recipient, which announce the presence of that dative argument. These morphemes are used to turn a (mono-)transitive verb into a ditransitive one (*dakart* 'I'm bringing it' → *dakarkit* '(s)he's bringing it to me'), although, as Schuchardt (1972 [1893]: 276) points out, in the western dialect we can find forms without *-ki-* which are ambiguous as to the two uses. Moreover, he mentions forms from Azkue (apud Schuchardt 1972 [1893]: 286) with a *-ts-* morpheme of the same meaning (*daroat* 'I'm carrying it' → *daroast* '(s)he is carrying it for me').

We should also mention that the verb *eduki* 'to keep, to have' was formerly a three-person variant of the verb **edun* (+ *-ki-*), lexicalized with its own monotransitive meaning. In the synthetic conjugation of this verb, there are two traces of its ditransitive origin: the DF *-k-* (< *-ki*), and the *-a-* of the 3rd person (*-o-* in the western dialect). Thus, *daukat* (B. *dekot*) would mean 'I have it to him/her', and later on it would acquire the meaning 'I keep it (for him/her)', losing the dative notion. Trask (*HB*: 228) cites a clause which could reflect an intermediate stage of that change: *edukiko dio bere emazteari* 'he holds his wife', lit. 'he holds (it) to his wife'.

The DF *-i-* also appears, although not in all dialects, in the verb *eman* 'to give' (*demaiot* 'I give it to him/her'), and in the auxiliary **ezan*. However, in the latter, depending on the dialect, it can appear just before the dative agreement or preceding the verb root, or even repeated in both positions (as in the standard

forms): *egin ezazu* '(you) do it!'→ *egin iezaiozu/ezaiozu/iezozu* (along with *ezozu*, without any DF) '(you) do it to him/her!'

As already discussed, Schuchardt related the markers *-ts-* and *-ki-* to each other as allomorphs, considering how close the verbs *e(d)uki* and *eutsi* appear to be to one another, but Trask (1995b) argued that this relationship was impossible, and Lakarra (2006a) proposed different etymologies for the two of them. Thus, *-ki* (which has an allomorph *-gi* in non-finite forms) would be related to the suffix *-ki* of the NP, and both would come from an old prefix which Lakarra himself (1995) detected in words such as *gizon* 'man', *gibel* 'liver', *gihar* 'muscle' or *gizen* 'fat'. Regarding *-ts-*, it would originally have been an allomorph of the suffix **-tz*, which, after gaining autonomy, replaced the original one. The old morpheme can be seen, according to Lakarra, in words like *ihintz* 'dew' (< **(n)i-nin-tz*), *hortz* 'tooth' (< *hor-tz* 'canine-*tz*', cf. Sp. *caninos* 'teeth'), *beltz* 'black' (< **bel-tz*, cf. *orbel* 'fallen leaves', lit. '~ dark leaves'), or in verbs like *ahantzi* 'to forget' (< **(n)a-nan-tz-i*). The *-ts-* variant probably appeared through analogy with the verbal noun, where **-tz-* would be between two consonants, hence turning into *-s-* and later on being restored as an affricate in the form of the perfect participle: *jauntz-i* [dress.PTCP] → *jaun(t)z-te* [dress.NOMNL] > *jaunste* → *jaunts-i*; thus, from **edun* 'to have' → *edu-ki* 'to hold' and **edun-tz-i* (> *eutsi* 'to hold to') (cf. *FHV*: 285-6, on *intsaus**st**i* 'walnut-tree grove' < **intzaurzti* < **intzaur-tz(e)-di*).

9.5.4.1.2. Alternation of the present tense vowel

In a certain old period of the language, the *a/e* alternation of the present tense vowel might also have been productive to mark the difference between the Nor-Nork and the Nor-Nori-Nork valency (Lakarra 1986), especially where the *dative flag* was not used. However, following Schuchardt (1972 [1893: 25]), we can assume that this alternation may be related to the aforementioned insertion of the *-i-* before the root, which could have had an influence on the *-a-* of the present, merging both in *-e-*. We have examples of this phenomenon from the oldest texts of the westernmost dialect (18a), as well as in the easternmost one (18b), so it might well have been a general mechanism in the old language:

(18) a. *Takok d**e**roat dodana neure etxerean zejara* Cf. *daroat* 'I take it to...'
'The need takes what I've got from my house to the market' (RS, 82)

b. *Ezakusan begik nigar ezt**e**gik (= ez degik)* Cf. *ez dagik* 'thou (M) will not...'
'The eye that cannot see thee (M) will not mourn for thee (M)' (Oih., *Prov.* 153)

9.5.4.2. Causativization as dative insertion

At this point, we should return to the originally causativizer prefix *-ra-*. It was already discussed in the subsection on the Nor-Nork type verbs, and now I can add that, by means of this morpheme, some ditransitive verbs have been derived from transitive ones. Examples include the following: the auxiliaries **eradun* (from **edun*; cf. the western equivalent *eutsi*) and **erazan* (from **ezan*); and the autonomous *erakarri* 'to attract' (from *ekarri* 'to bring'), *erakutsi* 'to show' (from *ekusi* 'to see'), *irakatsi* 'to teach' (from *ikasi* 'to learn'), and *eragon* 'to get busy with' (perhaps from *egon* 'to stay, to remain'), although the latter accepts only two arguments (in ergative and dative), despite having a ditransitive morphology (*ni-k lana-ri daragoi-o-t* 'I'm working, I'm busy with work').

9.5.5. More on the origin of the diathesis in the Basque verb: was Basque a passive language?

Beginning with Schuchardt (1972 [1893: 2]),[45] and continuing with Lafon (1980 [1944], 1999 [1954b])[46] up to Trask (1977),[47] many authors have argued for a passive nature (Schuchardt, Lafon) or origin (Trask) of the Basque transitive (finite) verb. At the roots of this hypothesis lies the interpretation (not to say the

45 "Das transitive Zeitwort des Baskischen ist passiv gedacht. *Na-kar-zu*, 'Sie tragen mich' heisst eigentlich 'ich werde getragen von Ihnen'". Although in a later published work, without necessarily contradicting his words, he takes the issue further when he says that "[e]l verbo [vasco] desnudo es indiferente, es decir, ni activo ni pasivo; por medio de los vocablos que le acompañan se precisará si está tomado en este o en aquel sentido" (1947 [1923]: 31; cited from the Spanish translation).

46 This author also mentions Friedrich Müller, to whom he attributes the authorship of the passive conception of the Basque transitive verb, as well as Stempf and a study by Gavel and Lacombe (Lafon 1980 [1944]: 15).

47 Among other things, Trask suggests that we should analyze the *-n-* of some non-present forms (e.g. *ni**n**duen* '(s)he had me' vs. *nuen* 'I had it') as a passivizer, even though he admits (and at the same time tries to refute) many objections, such as the fact that there are also "active" forms (plurals such as *ge**n**uen* 'we had it', in which it would be explained through analogy) and intransitive forms (e.g. *ni**n**tzen* 'I was', *ge**n**tozen* 'we were coming'), where the presence of such a morpheme would not be expected (Trask 1977: 210-212).

On the other hand, Trask applied the term *mediopassive* (*HB*: 112) to some instances such as *Sagarrak bildu dira* [apples gather.PTCP aux.INTR] 'Apples have been gathered', but as Martínez-Areta (2012) explains, these are only impersonal uses of the intransitive auxiliary. Besides, we should add that this type of clause is only possible in the periphrastic conjugation.

literal translation) of the agreement markers according to their relative position in the verb and comparing them to the NPs which they refer to (that is, according to the ergative-absolutive system). So, if *dator* is translated as 'it is coming', and *dakar* as 'it is bringing it', then which one of these last *it*s does that of *dator* make reference to? This is not a big problem with the 3rd person singular (for it is unmarked) but, when comparing forms which contain other persons, some scholars have resorted to a passive interpretation: *dator* '(s)he is coming', *dakar* '(s)he$_{1\text{-}2}$ is bringing him/her$_{2\text{-}1}$', *dakar-**t*** '(s)he is being brought **by me**'. Likewise, if ***n**ator* means '**I** am coming', then ***n**akar* means '**I** am being brought by him/her'. This, added to the fact that the ergative-absolutive case morphology can have similar interpretations, has led to a great interest in the topic, but as yet no satisfactory diachronic explanation has been reached, since a diathetical system in which only the passive voice exists is not easily conceivable, and there is no trace whatsoever in the Basque language which could lead us to think that there was ever a voice opposition *à la indoeuropéenne*.

Recently, Martínez-Areta (2012) has analyzed the possible origin and evolution of the Basque diathetical differences based partly on the "passivist" theses, especially that of Lafon (1999 [1954b]), although Martínez-Areta's conclusions are rather different, more detailed, and more interesting.[48] According to this linguist, forms such as *nakar* would not originally have been passive (or, more precisely, they have never been exclusively so), but Proto-Basque was a language without a grammaticalized diathesis, similar to Chinese, in which verbs would acquire their diathetical value depending on the surrounding elements.[49] That is to say, *nakar* would originally mean both 'I am bringing (him/her)' and 'I am being brought (by him/her)', and the ambiguity would be resolved by the context. Martínez-Areta proposes some disambiguating mechanisms (word order in b_1-c_1, agentive complement in b_2-c_2, in Figures 9.3 and 9.4). Figure 9.3 illustrates this process with the verb *ikusi* 'to see' in cases in which both A and P are 3rd p.

From the options in b-c of that figure, the ones with a passive interpretation would generalize (for some unknown reason),[50] and, after the other disambiguat-

48 Among other things, he tries to show the consequences that are derived from his hypothesis for other areas of reconstruction such as word order, and whether Proto-Basque was a *topic-prominent* or a *subject-prominent* language (cf. §10.4.4).

49 Cf. modern Mandarin Chinese: *yú chī-le* [fish eat-PFV] 'The fish has eaten ~ The fish has been eaten' (apud Martínez-Areta 2012).

50 This is probably the weakest point of the explanation, although when discussing the non-present forms, Martínez-Areta cites an interesting quote from Trask: "In Yana, a Hokan language [...], the passive is obligatory whenever the patient is first or second person" (apud Martínez-Areta 2012). Through that, he explains the *split* of the non-present forms,

ing options had been lost, they would be reinterpreted as active, leading to an ergative system.

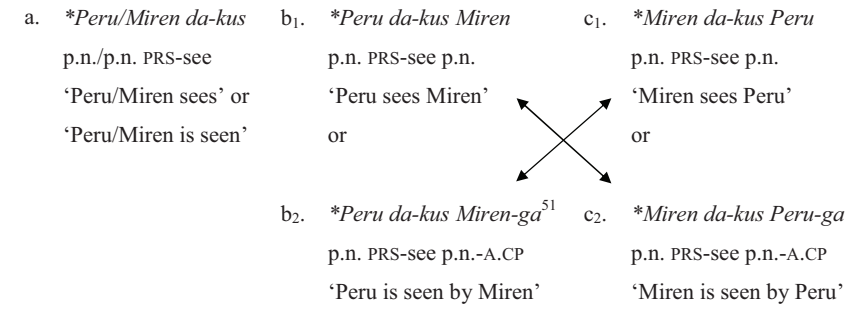


Figure 9.3. Proto-Basque diathetical system of the present tense according to Martínez-Areta (2012)

Things become even more complicate if we consider the *split* of the non-present tenses, whereby, depending on whether the direct object is a 3rd person or a 1st-2nd person, the agreement morphemes change with regard to their present tense equivalents (***n**au* ~ ***n**induen*, but *du**t*** ~ ***n**uen*). To explain this difference, Martínez-Areta proposes that, in the cases in which the direct object was the 3rd person, the prevailing interpretation was the active one, as in b_1 of Figure 9.4, not the passive one as in the rest of the cases, e.g. b_2 of the same Figure. Thus, in some cases, two current forms have emerged from just one proto-form. This is illustrated in Figure 9.4 with the verb **edun* 'to have'.[52]

According to this theory, through a change which occurred due to morpho-phonological processes in the final *-n* of the pronominal bound morpheme, and the adaptation to the new ergative-absolutive system, the two forms which we find historically —that is, *nuen* 'I had it' and *ninduen* 'it had me'— were both derived from a single proto-form, given in b_1 and b_2 of Figure 9.4: **nen-dun-en*.

---

but it could also be applied to the present tense, since the 3rd person has no agreement marker and it is not possible (in other words, it is irrelevant for us) to know if *dakus* is to be understood in its origins as '(s)he sees him/her' or as '(s)he is seen by him/her'.

51 This *agentive complement* marker *-ga* could be the origin of the ergative suffix *-k* (cf. §9.3.1 on the NP pluralizer and 2nd p. sg. agreement marker **-ga* > *-k*).

52 I have simplified the explanations of these examples by glossing **nen-* as [1SG.NPRS] (the same can be said for **hen-* [2SG.NPRS], **gen-* [1PL.NPRS], and **zen-* [2PL.NPRS]). I shall take up the whole issue concerning these elements in more depth in §9.8.

As we can see, this is a very interesting path for further research, and it could be crucial for a better understanding and reconstruction of the Proto-Basque finite verb.

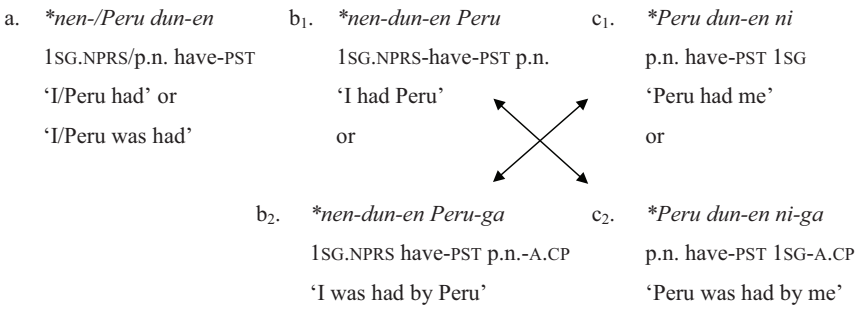


Figure 9.4. Proto-Basque diathetical system of non-present tenses according to Martínez-Areta (2012)

9.6. Allocutive agreement

In this section, I discuss the allocutive verb forms in Basque, both with regards to their use and to morpho-phonological issues; I focus on the standard variety of the language, but, where necessary, I also show the main dialectal differences.

It is quite common in languages to find the insertion of arguments (usually in the dative case) which do not directly participate in the action expressed by the verb, but which are affected in some way or to some degree by it. This has been used to different extents (but is usually not compulsory) in different languages, including both old ones such as classical Greek or Latin and modern ones such as German, Spanish, and even English. In these languages, especially in pronouns, the so-called ethical dative (*dativus ethicus*) and/or the benefit ~ harm dative (*dativus (in)commodi*) are used.

In his great *Baskische Studien* (1972 [1893: 10-11]), the Austrian linguist Schuchardt explained that traditional Basque has known many allocutive systems, both in the familiar register and in the polite or neutral one, for all persons. Therefore, to ask 'who are you?' people often say *nor zaitut/zaitugu?* (lit. 'who do I/we have you?'), and it is very common to say, for example, *semea dut* (lit. 'I have him (as) a son') to express 'he is my son'. Schuchardt also notes the closeness between the transitive and the bivalent intransitive forms to express the idea of implication/affectation, e.g.: *neke handia duzu*, lit. 'you have it as a

great fatigue' : *neke handia zaizu*, lit. 'it is a great fatigue to you', both meaning 'you find it very tiring'.

In these uses, Schuchardt sees the origin of the Basque allocutive system: a complete and regular system of allocutive forms for all verbs, in which the addressee (that is, the 2nd person singular) appears indexed, sometimes as a dative (19), sometimes as an agent (20), and in the rest of the cases as a *second dative* (21-22). As in the cases of the aforementioned special datives, these arguments are not directly involved in the action. Besides, it is not necessary that they should be interested in it or feel affected by it in any way. In this sense, in the dialects where polite allocutive forms are used, the above mentioned clause *neke handia duzu hori* can also mean 'it is very tiring', without involving any affectedness as far as the second person is concerned.

In the following examples, I give a list of allocutive verbs, compared to their non-allocutive equivalents with regard to their meaning:[53]

(19) *ø-d-i-a-t* vs. non-alloc. *dut*
3-PRS-have.to-2SG.ALF.M-1SG
'I have it, man'

(20) *n-a-u-n* vs. non-alloc. *naiz*
1SG-PRS-have-2SG.ALF.F
'I am, woman'

(21) *ø-za-i-gu-k* vs. non-alloc. *zaigu*
3-be.PRS-DF-1PL-2SG.ALF.M
'(S)he is to us, man'

(22) *ø-za-kar-ki-o-na-gu* vs. non-alloc. *dakarkiogu*
3-PRS.AL-bring-DF-3-2SG.ALF.F-1PL
'We are bringing it to him/her, woman'

It is important to note that allocutive forms of the Nor and Nor-Nork types can be identical with non-allocutive forms of other types (Nor-Nork and Nor-Nori-Nork, respectively). For instance, the *diat* of (19) is allocutive, but *diat* is

53 All of the examples given here belong to the form of address *hika* (familiar register) which corresponds to the original 2nd person singular (*hi*); there also exists (but dialectally more limited) the allocutive use of *zuka*, that is, the new 2nd person singular (*zu*, former plural, cf. §9.3), as well as its expressive or affective variant *xuka*, which is used in an even more restricted geographical area (cf. §9.6.3).

also the non-allocutive form of the tripersonal auxiliary, meaning 'I have it to thou (M)'. Similarly, the *naun* of (20) can also mean 'thou (F) hast me' (non-allocutive).

9.6.1. The use of allocutive forms

Allocutive forms usually appear as a way to catch the listener's attention and involve him/her in the conversation, but in the Basque language they have been grammaticalized as compulsory (within certain limits), so that a clause can be ungrammatical if one of these forms is not used where it should be, or if it is used where it should not (although these restrictions tend to relax).

9.6.1.1. Morphosyntactic triggering and constraints

Forms such as those given in the previous examples are obligatory whenever the speaker addresses the listener as *hi* 'thou' (and in some dialects the same occurs with *zu* 'you (sg.)', cf. §9.6.3), but some morphosyntactic environments prevent its use:[54]

a) Interrogative and exclamative clauses:

(23) a. *Zer du (**dik/n) horrek?*
'What's wrong with him/her?'

b. *Itsusia da (**duk/n) gero!*
'It is ugly, for sure!'

b) Subordinate clauses built with *-(e)n*, *-(e)la*, *bait-*, *ba-* (in its conditional sense 'if'):

(24) a. *Esan dik/n berdin zaiola (**zaio(n)ala)*
'(S)he says that (s)he doesn't care'

b. *Gaur autoa ez badakarte (**bazakartek/n), ezingo nauk/n etxera joan*
'If they don't return my car today, I can't go home'

54 Here I show the system which is common to all the dialects and is the one reflected from the oldest texts (Lafon 1980 [1944]: 405-407), and which has been adopted for the standard variety. For a more current perspective of the spoken language, cf. Aurrekoetxea (1994: 996-7, 1000-1), where the use of allocutive forms in interrogative clauses and in some subordinate clauses is shown for the westernmost varieties.

c) Subjunctive forms:[55]

(25) a. *Haien bila goazen (**goaza(n)an) eskatu ziguk/n*
'(S)he asked us to go to pick them up'

b. *Lana hobeto dagiten (**zagite(n)an) nahi di(n)agu*
'We want them to do the work better'

d) 3rd person of the imperative:[56]

(26) a. *Betor! (**Betorrek/n!)*
'Let him/her come!'

b. *Dakarrela! (**Zakarre(n)ala!)*
'Let him/her bring it!'

To sum up, although the use of the allocutive forms is obligatory for the main verbs of the clauses, they cannot be used in non-declarative ones.[57]

What I have said up to now only refers to the grammatical features of this system. In the following subsection, I extend my analysis to more social and pragmatic factors.

9.6.1.2. Dialectal extension[58]

The use of the pronoun *hi* and its corresponding verb inflection has decreased in the course of history, more quickly or slowly depending on the dialect, but we can say that the greatest loss has occurred during the last few decades. This has occurred for various reasons: firstly, because of social taboos as to its use (it has been considered rude or lacking respect), and secondly, due to the recent literacy process in which this register has not been fully taken into consideration. As a

55 It could not be formed in another way because the subjunctive forms are mere subordinate forms with *-(e)n*.

56 This restriction is probably related to the first two, at least in cases like (26b).

57 "Es ist eine allbekannte, von Inchauspe, dem Prinzen Bonaparte u. A. deutlich ausgesprochene, von van Eys TB 24. 27 nur in etwas unsicherer und in unvollständiger Weise vorgebrachte Thatsache dass in Frage-, Heische- und untergeordneten Sätzen statt der verschiedenen Bezugsformen die Gemeinformen angewendet werden. [...] [S]o muss man sagen dass der Redende durch grammatische Gründe bestimmt wird zwischen der dutzenden und der höflichen Anrede zu wechseln" (Schuchardt 1972 [1893: 9]).

58 In this subsection and that of §9.6.3, I am following the articles on this issue by Jabier Alberdi (1995, 2011), which are in turn based on his doctoral dissertation.

result, younger generations are unable to speak fluently in *hika*, either because they did not learn Basque at home, or because in their social environment this register was no longer used. Therefore, it is possible that what the future holds for this pronoun (and hence for the allocutive system) is the same as what has happened to the English *thou*.

Nevertheless, as previously discussed, not all dialects have lost the use of this form, and in some of them it is still widely used. However, there are sometimes differences between men and women speakers: for instance, there are cases of men using it and women not doing so, or both men and women using it but only when addressing men.

This tendency for the 2nd person singular pronoun of familiar register to disappear may have been brought about by the use of the 2nd person (originally) plural as a singular of respect, which eventually became increasingly less marked, until it took over the whole 2nd p. sg. slot. To make up for the semantic change of *zu* 'you (pl. → sg.)', a new 2nd plural (*zuek*) came about, and in the dialects in which the loss of *hi* was most advanced (westernmost varieties), a new polite register *berori* (< **ber hori* 'that very') emerged, which agrees with the 3rd person (cf. Spanish *usted*). However, this has fallen into disuse among the young generations (cf. §7.3.1).

Besides these uses, in the eastern dialects there are two other allocutive systems: *zuketa* and *xuketa*, corresponding to *zu* 'you' and *xu* 'you (affective)'. In these systems, an allocutive morpheme has to be inserted even in the cases where *zu* or *xu* is not an argument of the clause, as occurs with the *hiketa* of the rest of the dialects. The first system is used (or was used, since some of these varieties have already disappeared) in Soule, in the eastern part of Low Navarre and in the valleys of Salazar and Roncal in High Navarre. The second system is more limited to the northern part of these territories, that is, to Low Navarre and Salazar (cf. §9.6.3).

9.6.2. The hiketa

Up to now, I have mainly discussed the allocutivity which corresponds to the pronoun *hi*, but I have also mentioned the one corresponding to *zu* and to its palatalized affective variant *xu*. Now, I shall discuss the *hiketa* system in more detail, since it is the one which is general to all the dialects, and the only morphological difference between the two registers is, obviously, the difference in the agreement marker. Then (in §9.6.3), I shall present more data on the allocutivity of *zu* and *xu*.

In this subsection, I shall discuss the morphological and phonological changes derived from the allocutive agreement of the familiar register (*hiketa*), which are produced in the conjugated verb. As Lafon (1980 [1944]: 407) emphasized, even though similar mechanisms are used in many languages (such as the use of dative forms, or the use of the verb *to have* instead of the verb *to be*), the originality of the Basque language lies in the systematic and regular use of allocutivity.[59] Moreover, Lafon highlights the fact that the dative forms and the allocutive ones are not exactly the same (unlike in French, for instance), and that they distinguish the listener's sex.

9.6.2.1. Morphological changes

We can divide the allocutive forms of the Basque verb into three groups following Schuchardt's (1972 [1893]) classification. Firstly, we have the forms which the Austrian linguist called "proper forms" of the allocutive, that is, the ones in which the agreement marker is simply added to the non-allocutive verb form (or at most they change also the initial consonant), as in (27a) and (27c). Secondly, there are some others which he calls "non-proper", because they change the root (§9.6.2.1.2) and resemble another non-allocutive form with an ergative or dative 2nd person agreement (27b). And thirdly, we have the *mixed* forms, which change their root in the same way as the "non-proper" ones, but this time they identify with non-allocutive forms without the 2nd person agreement, to which that marker is simply added in the same way as in the "proper" ones (e.g., *litzateke* 'it would be' → *lukek/n* 'it would be, man/woman', which recalls *luke* 'it would have it' + *-k/n*).

9.6.2.1.1. Addition of the addressee agreement

As I said in the introduction of this section, the main morphological feature in the allocutive system is the addition of the listener's agreement marker (2nd p. sg.). This usually takes the dative agreement form (27b), but sometimes the ergative one (*duk/n* 'it is, man/woman' ~ 'thou [M/F] hast it').

59 It is usually summarized by saying that when the speaker addresses the interlocutor with the *hi* pronoun, (s)he must use (always within some morphosyntactic limits) the corresponding allocutive verb forms. This is true from the earliest documentation (Dechepare 1545, Leiçarraga 1571), but the text *Refranes y Sentencias* (1596) and Lazarraga (ca. 1607) provide us with some examples of a previous stage in which such an obligation did not exist (Kintana 2005: 155 ff.). Cf. §9.6.3 on the polite form of address.

We must note that, when there is another dative or ergative agreement in the verb, the allocutive one appears in a particular position, that is, behind the dative agreement and before the ergative one:

(27) a. *z-i-o-(n)a-t* (allocutive form derived from neuter *di-o-t*)
PRS.ALF-have.for-3.**DAT**-2SG.**ALF**-1SG.**ERG**
'I have it to him/her, (wo)man'

b. *d-i-(n)a-t* (allocutive form derived from neuter *du-t*)
PRS-have.for-2SG.**ALF**-1SG.**ERG**
'I have it, (wo)man'

c. *za-i-da-k/n* (allocutive form derived from neuter *zai-t*)
be.PRS-DF-1SG.**DAT**-2SG.**ALF**
'(S)he is to me, (wo)man'

9.6.2.1.2. Stem suppletion

Another important morphological change, which in any case affects only two verbs (*izan* 'to be' and **edun* 'to have', both used as auxiliary and synthetic verbs) is the change or suppletion of the stem; that is, the use of different (etymologically unrelated) roots according to paradigmatic differences. The semantically closest verb which contains an extra argument is used, thus changing from the verb *to be* (one argument) to the verb *to have* (two arguments) (28), and from *to have* to *to have for* or *give* (three arguments) (29). As a result, the allocutive forms corresponding to *izan* are exactly the same as the non-allocutive forms of **edun*, and some allocutive forms corresponding to **edun* are the same as the non-allocutive forms of *-i-*. These changes are common to all dialects, except for the western one, where **edun* itself is used for allocutive forms of **edun* (cf. §9.8.1):

(28) *izan* → **edun*:
a. *da-**ø*** '(s)he is' → *d-**u**-k/n* '(s)he is, man/woman' or 'thou (M/F) hast it'
b. *n-a-**iz*** 'I am' → *n-a-**u**-k/n* 'I am, man/woman' or 'thou (M/F) hast me'

(29) **edun* → *-i-*:
a. *d-**u*** '(s)he has it' → *d-**i**-k/n* '(s)he has it, m./w.' or '(s)he has it to thee'
b. *n-a-**u*** '(s)he has me' → *n-a-**i**-k/n* '(s)he has me, m./w.' vs. *d-i-t* '(s)he has it to me'

As we can see, in the examples given in (28) the 2nd person index does not belong to a dative argument but to an ergative one, which would seem to be an even greater argument promotion, since the ergative is the subject of the transitive clause. Apparently, the underlying reason is the mixed formation of the allocutive forms with regards to their origin and composition. While in the great majority of the verbs and forms the addition of an allocutive argument does not affect the root, and only the new morpheme is added (and sometimes the *dative flag* *-i-*, cf. §9.6.2.2), in the aforementioned cases the argument structure of the verb changes because the original root is replaced by a "more transitive" one. For this reason, from *da*, *naiz*, etc., instead of the expected changes *da* → *zaik*, *naiz* → *natzaik*[60] (as with *dago*, *nago* → *zagok*, *nagok*) we obtain *duk*, *nauk*. As far as the change *du*, *nau* → *dik*, *naik* is concerned, many specialists have seen in it a reduced form of the same root (**edun* 'to have') plus the *dative flag* (Schuchardt 1947 [1923]: 79-80; for the non-allocutive forms such as *dio* 1972 [1893]: 289, and Gómez & Sainz 1995: 239, cf. §9.2.1). For example:

(30) *dik*, *naik* < **d-u-**i**-**k**/*d-**i**-u-**k**, *n-a-u-**i**-**k**/*n-a-**i**-u-**k** ← *d-u*, *n-a-u*

However, there are two problems with this hypothesis. Firstly, if we choose the **-ui-* variants, this would imply that the *dative flag* had been placed after the verbal root of these forms, going against all the other cases of the allocutive system in which it appears before the verbal root. If we accepted the reconstruction of forms with **-iu-*, we would lack general parallels for the *-iu-* > *-i-* reduction, and besides, we would leave unexplained the absence of the change *d-* > *z-* (cf. *da* → *duk*, ***zuk*, where no dative flag has ever been inserted, and instead a replacement of the root has taken place; see §9.6.2.2).

Lakarra (2008a) has recently proposed identifying these forms (along with those belonging to the ditransitive auxiliary) with the old verb root *-i(n)-* 'to give' (which he reconstructs as **nin*), since it is very common among the languages of the world to use this verb to introduce the goal/beneficiary argument in ditransitive constructions. There are two main advantages to this hypothesis: 1) it explains the vocalism of the root (common to all the dialects except the western one, at least nowadays),[61] and 2) it indirectly accounts for the

60 Indeed, these forms exist, but their meaning is 'it is to thou (M)' and 'I am to thou (M)', respectively.

61 In the western dialect, the pair *du/dik* is rendered *dau/jok* (the second form also being from **edun*). The latter form can be reconstructed as < **da-u-k* plus the change in the initial consonant either caused by the *dative flag* or because of its analogy with some other allocutive forms. However, the forms that belong to the *-i-* root are also documented in

lack of the *d-* > *z-* change in the allocutive forms of *du* (*dik*, not ***zik*), since, originally being a three-person verb, it does not need the *dative flag* to add the agreement that belongs to the allocutive argument. It is true, anyway, that there exist forms of the 3rd p. with *z-* (probably due to analogical pressure) found in a somewhat peculiar dialectal distribution. In the language of Leiçarraga, who usually wrote forms with *dia-* (not *za-*), we find (along with *ditik*) the form *zitik* for the plural of *dik* (the allocutive forms of *ditu* and *du* respectively), which is similar to the *zetik* of some local Guipuscoan varieties. In the reverse cases, in the varieties in which forms with *z-* are usually used, we will nearly always find forms with *d-* for this verb.

9.6.2.2. Phonological changes

I have already mentioned some phonological features involved in the construction of allocutive verbs. Now, we shall look more closely into these and some other changes which have occurred due to the addition of the morpheme *-i-*, which —as Schuchardt (1972 [1893]) observed—[62] we should directly or indirectly identify with the *dative flag* (cf. §9.5.2 and §9.5.4.1.1).

9.6.2.2.1. Initial d- *and the* -i- *insertion*

Even though I am conscious that it would be more accurate to describe what I have called "the *-i-* insertion" in the section on morphological changes, I believe it is more useful to discuss it here (and in §9.6.2.2.2), considering the phonological changes which it seems to have brought about.

To begin with, we need to bring up the *dative flag* again, already discussed in §9.5.3 and §9.5.4.1. It is quite certain that we are talking about the same morphological element in both cases, even though the one that concerns us now does not show any other allomorph, unlike the previous one (*-ki-*),[63] and does not mark the presence of a dative, but instead represents the listener's agreement marker. Another difference is that the *dative flag* appears both prefixed and suf-

---

the earliest texts of the western varieties (which is a further piece of evidence against considering the root *-i-* to be a reduced form of **edun* + *dative flag*).

62 "Nachdem ich das "i" der Zielformen so ausführlich besprochen habe, will ich zeigen dass das "i" der Bezugsformen mit ihm identisch ist" [1893: 59], cf. eastern/western *diagok/jagok* 'it is thy (M)' and 'it is, man', although in some varieties (as well as in the standard language) there is a tendency to distinguish between the two meanings: *dagokik* 'it is thy (M)' vs. *zagok* 'it is, man' (cf. the following footnote).

63 In some varieties and in the standard language, in fact, the variant *-ki-* is almost always used for the real dative.

fixed, whereas in the allocutive forms this *-i-* always appears (or its presence can be deduced) in a position before the verbal root. This would assign it some antiquity, although it could also be a specialization of the oldest form for the "modern"[64] (more restricted) use, following Kuryłowicz's famous law. In any case, the allocutive argument usually appears as if it were a dative. Hence, it is not surprising that the same mechanism of adding arguments should appear in both cases. The dative would grammaticalize first, then some particular uses of it (such as the *dativus ethicus*) would spread, along with some other mechanisms of the listener's "argument promotion", and finally the allocutive system would be regularized.

Turning to the heading of this subsection, in the majority of the dialects, the *-i-* has had an influence on its neighboring sounds in a somewhat exceptional way within the historical phonology of Basque (cf. *FHV*), probably due to a *glided pronunciation* which made the (semi-)vowel and the consonant merge.[65] However, it is not clear what the position of the vowel was, whether it was found before (**i-d-*) or after (**d-i-*) the consonant. In either case, the results are the following (Aurrekoetxea 1994: 1003): western dialects *j-* /j/, /ʤ/, /x/ (*jagok*), eastern dialects *di-* /dj/ (*diagok*), central varieties and standard *z-* /s̻/. In some local varieties, *x-* (/ʃ/) (*xagok*) is also found. Either because of conservationism or because they have innovated, there are some instances of *d-* without any glide.

Here are some examples of the standard variety, the morphology of which is based on the central dialects:

(31) N a. *dago → zagok/n*
'(S)he is' → '(S)he is, (wo)man'
Cf. *dagokik/n* 'It corresponds to thee'

64 Indeed, the dialectal variation found in the outcomes of this process leads us to believe that the allocutivity, as we historically know it, is a relatively modern phenomenon, that is, subsequent to the fragmentation of the common language (Alberdi 2011).

65 Here, we always talk about the initial *d-* consonant of the present forms with 3rd person agreement (*ø-*) as a subject or patient. In some dialects, certain other consonants such as the *n-* of the 1st person singular (**nj- > ñ-* /ɲ/) or the *l-* of the hypothetical tense (**lj- > ll-* /ʎ/) have also been altered. Some scholars (cf. Alberdi 1995: 284-285) have defended the idea that the palatalized forms such as *ñagok*, *jatork,* etc. are the result of the morphologization of phonological facts (meaningful palatalization), and that the *-i-* of variants such as *niagok*, *diatork* appeared through depalatalization. Even if this were the case, the expressive palatalization does not account for the *z-* of the central varieties, so we should perhaps consider two interlaced processes.

b. *dator → zatorrek/n*
'(S)he is coming' → '(S)he is coming, (wo)man'

(32) NN_1 a. *darrait → zarraidak/n*
'(S)he is following me' → '(S)he is following me, (wo)man'

b. *dagokit → zagokidak/n*
'It corresponds to me' → 'It corresponds to me, (wo)man'

(33) NN_2 a. *daramagu → zarama(n)agu*
'We are carrying it' → 'We are carrying it, (wo)man'

b. *dakargu → zakarre(n)agu*
'We are bringing it' → We are bringing it, (wo)man'

(34) NNN a. *daramakigu → zaramakiguk/n*
'(S)he is carrying it for us' → '(S)he is carrying it for us, (wo)man'
Cf. *daramakiguk/n* 'Thou are taking it for us'

b. *dakarkigu → zakarkiguk/n*
'(S)he is bringing it to us' → '(S)he is bringing it to us, man/woman'
Cf. *dakarkiguk/n* 'Thou are bringing it to us'

9.6.2.2.2. Tense vowel and the -i- *insertion*

In the central varieties, this *-i-* does not only modify the consonant, but —as I already mentioned in §9.5.4.1.2 when discussing the old vowel alternation between two- and three-person verbs— it also alters the following *-a-* of the present tense, and as a result we have an *-e-* (cf. Aurrekoetxea 1994: 1004). Compare the following examples with the corresponding ones given above (31-34):

(35) *zetorrek/n*, *zegokidak/n*, *zerama(n)agu*, *zekarkiguk/n*

In the western varieties, the *-i-* has affected only the consonant, whereas, in the eastern ones, it has been kept in the majority of the cases without any alteration of the following vowel.

9.6.3. The zuketa *and* xuketa

To round off the section on allocutivity in the Basque verb, I shall now focus on the *zuka* and *xuka* forms of address, which belong to the use of the pronoun *zu* 'you (sg.)' and to its palatal affective variant *xu* 'id.', some details of which have already been discussed.

The connection between these two forms of address is asymmetrical, as far as their dialectal area is concerned: in the places where *xuka* is used, *zuka* is also used, but not the other way round. The most widespread form (*zuka*) includes Souletin, Eastern Low Navarrese, Roncalese and Salazarese. The most reduced one (*xuka*), however, only appears in Low Navarrese and in Salazarese.

I have already pointed out that the only morphological differentiating device between the various allocutive systems of the 2nd person singular (familiar *hi*, polite *zu*) is their agreement marker (*-(n)a-*, *-k/n* for the first one, *-zu(-)* for the second one). Therefore, it is exactly the same difference which is found between the ergative or dative markers of these persons. Hence, in the *zuka* system (as well as in the *xuka* one), no differentiation is made as to the listener's sex.

As for their use, I should make it clear that, although the pronoun *zu* (for the 2nd p. sg.) is used in the formal/neuter register, the *zuka* allocutive also implies a closeness or familiarity in the dialects in which it is used, as opposed to the non-allocutive use of the *zuka*,[66] which is reserved as the more formal and distant register. At the same time, it is opposed to *hika* because this is more informal than the *zuka*. In the dialects in which there is no allocutive *zuka*, the neuter takes its place (and, for many speakers, it even takes the place of *hika*, in this case being the non-allocutive *zuka* like the English *you*). For a more formal register, historically the aforementioned *berorika* appears, in which the verb shows 3rd person agreement (cf. the *usted* form of address of Spanish). This is illustrated in Table 9.11.

Table 9.11. Forms of address in dialects with and without allocutive zuka

	Dialects without alloc. *zuka*	Dialects with alloc. *zuka*
Formal-polite	*berorika*	non-allocutive *zuka*
Formal-familiar	non-allocutive *zuka*	allocutive *zuka*
Informal-familiar	*hika*	*hika*

To all these forms of address, we should add the eastern allocutive *xuka*, which would be positioned between the *hika* and the *zuka*, since it is a palatal-

66 It is worth recalling that the *hika* shown in the previous sections does not have a non-allocutive use, at least nowadays (cf. §9.6.1.1, in footnote).

ized expressive variant which shows greater affection or closeness than the allocutive *zuka*, but without the informal nuance of the *hika*.

With regard to the evolution of the allocutivity of the *zuka*, it should be pointed out that this has not occurred at exactly the same time nor to the same degree as that of the *hika*. While allocutivity is obligatory within the *hika*, in the *zuka* there is an allocutive/non-allocutive alternation (Lafon 1980 [1944]: 408-409) from the earlier texts (Dechepare, Oihenart);[67] an alternation which in any case does not seem to show any register change:

(36) a. *Amore bat ukhen **dizit** miragarri jentilik / harekila ninzanian e**nuien** nik faltarik* (Dch. VII, 3-4)
'I have had a love who is wonderfully charming / When with her, I lacked nothing'

b. *Ur-sortak, ardur'arduratuz / harria xila **zirozu*** (Oih., *Poems* I, 6th stanza)
'A constant dripping of water / can bore the stone'

The example (36a) is taken from Dechepare (1545), and there we can clearly see the alternation between the allocutive (*dizit*, neuter *dut* 'I have') and the non-allocutive (*nuien* 'I had', allocutive *nizin*) forms. (36b) is by Oihenart (1657), and, even though the allocutive form (*zirozu*,[68] neuter *diro* 'it can') does not alternate with any other non-allocutive one in the same clause, some lines above in the same poem, we find neuter forms such as *dut* 'I have' and *da* 'I am', instead of the allocutive ones *dizut* ~ *dizit* and *duzu*.

Later, the use of allocutive forms with the *zuka* form of address became regularized as those of *hika* had done, and analogically, the variant *xuka* was created.

Therefore, we can conclude that, unlike the common *hika*, the allocutivity of the *zuka* fully developed in one part of the Basque territory alone. As the examples from Zavala and Lardizabal suggest (apud Schuchardt 1972 [1893: 59]), the use of such forms (= *ethical dative*) of the *zuka* may have been more widespread than they are nowadays, even if they never became established outside the aforementioned area. Lardizabal (ibid.) cites the following example from a central variety:

67 Although Oihenart was from Soule, his writings reflect a language that was in some aspects similar to the Labourdin dialect, which lacks the allocutive *zuka*. But cf. the example (36b).

68 Note the initial *z-*, which corresponds to the initial *d-* of the non-allocutive form.

(37) *Nik jan **nizun** nere aitaren etxean*
'I ate at my father's house' (where ***nizun*** stands for the neuter *nuen* 'I had')

9.7. Auxiliary verbs and synthetic verbs

Throughout this paper, I have almost exclusively focused on the conjugated forms of the so-called *synthetic verbs*, that is, those which express the tense, mood, person and number categories, along with the lexical meaning, in a single form. The periphrastic constructions (i.e. those consisting of more than one element) have been described in Chapter 8. However, here I consider it necessary to tackle some issues about them, since they also contain conjugated forms (the auxiliary verbs). In the following paragraphs, I will show the main differences between the two systems, as well as their historical/genetic relationship.

9.7.1. Differences

9.7.1.1. Paradigmatic differences

The periphrastic conjugation contains a wider range of TAM distinctions than the synthetic one; however, this does not mean that the conjugated forms themselves are richer in one case than in the other. The key is that: a) since there is a rich root suppletion in the finite forms of periphrastic constructions (auxiliaries), their tense and mood distinctions are more "effective", and b) the possibilities of combination of the auxiliary verb with different forms of the lexical (non-finite) verb allow them to express more aspectual values. In Table 9.12, each column corresponds to a form of the main verb, and each row corresponds to a form of the auxiliary verb. For each possible combination, I have also included the semantically equivalent synthetic form (if there is one).

As we can see in that table, the periphrastic constructions can inevitably make more distinctions than the synthetic forms, which undergo a strong syncretism, since plenty of them have (or had in the past) more than one meaning. Without a doubt, this must have had an influence (among other factors) on the increasingly greater prevalence of the periphrastic forms, probably due to the universal preference for the overt marking of the central differences.

Table 9.12. TAM distinctions expressed by periphrastic combinations

		etor-ø	*etor-tzen*	*etorr-i*	*etorr-i-ko*
izan	*da*	- -	‘(s)he (usually) comes’ *dator* ‘(s)he is coming’	‘(s)he has come’ -	‘(s)he will come’ arc. *datorke*$_1$
	zen	- -	‘(s)he used to come’ *zetorren*$_1$ ‘(s)he was coming’	‘(s)he came’ arc. *zetorren*$_{1bis}$	‘(s)he would’ve come’ *zetorkeen*$_1$
	litzateke	- -	‘(s)he would come’ *letorke*$_1$	‘(s)he would’ve come’ -	‘(s)he would come’ *letorke*$_1$
**edin*	*dadin* [PRS]	‘that (s)he come’ arc. *datorren*	- -	- -	- -
	zedin [PST]	‘that (s)he come’ arc. *zetorren*$_2$	- -	- -	- -
	daiteke	‘(s)he can come’ arc. *datorke*$_2$	- -	- -	- -
	zitekeen [PST]	‘(s)he could come’ arc. *zetorkeen*$_2$	- -	- -	- -
	liteke [HYP]	‘(s)he could come’ arc. *letorke*$_2$	- -	- -	- -
	bedi	‘let him/her come!’ arc. *betor*	- -	- -	- -

As yet, it is not fully clear as to how the auxiliaries grammaticalized in their current functions (but now cf. Mounole’s doctoral dissertation, 2011). They not only differ in their roots, but also in the form(s) which the lexical verb can take when combining with them, and the result is a complementary distribution (*izan/*edun/-i-* with forms with overt aspect marking: *-tu* [PFV], *-tzen* [IPFV], *-ko* [PROSP]; **edin/*ezan* with forms which do not have any aspect marking).[69]

69 Besides, the diathetical structure of the verb is also distinguished by means of the auxiliaries: *izan* for Nor and Nor-Nori, **edun* for Nor-Nork, and *eutsi/-i-/*eradun* for Nor-Nori-

Therefore, we cannot simplify the issue by saying that each auxiliary has grammaticalized in order to express certain moods. Instead, each auxiliary —along with its corresponding combinatorial possibilities with the lexical verb— has grammaticalized or specialized under certain semantic-aspectual conditions.

9.7.1.2. Differences in meaning

Firstly, we should recall that nowadays there are few verbs which have a strong conjugation besides the periphrastic one. Moreover, they are all defective, although some of them have a rather broad synthetic paradigm. However, this does not mean that "binary" periphrastic conjugations are capable of expressing all the possible aspectual nuances. For example, in order to periphrastically express the progressive aspect which synthetic verbs can bear (among other meanings such as gnomic, planned future, stative with state verbs, etc., like its Spanish equivalent), we must add the particle *ari*[70] to the construction. Here are some examples of that aspectual contrast:

(38) a_i. *Etxera* *joa-ten* *naiz/nintzen*
to.home on-IPFV be.1SG.NPST/PST
'I usually go home / I used to go home'

a_{ii}. *Berandu* *iris-ten* *naiz/nintzen*
late arrive-IPFV be.1SG.NPST/PST
'I usually come late / I used to come late'

b_i. *Etxera* *noa/nindoan*
to.home go.1SG.NPST/PST
'I am/was going home'

b_{ii}. *Berandu* *iris-ten* *ari* *naiz/nintzen*
late arrive-IPFV PROG be.1SG.NPST/PST
'I am/was coming late'

Nork in the indicative mood; **edin* for Nor and Nor-Nori, **ezan/egin* for Nor-Nork, and the same ones (with the insertion of the *dative flag* in **ezan* and the tense vowel change *-a-* > *-e-* in *egin*) or **erazan* (< **ezan* + caus. *-ra-*) for Nor-Nori-Nork in non-indicative moods.

70 Probably a grammaticalization of **har-i* 'bussy, engaged' < lit. 'held', as proposed by Aldai (2003). Cf. the same root with the modern participle *har(-tu)* 'to take'.

As we can see, the sentences ($38a_i$) and ($38a_{ii}$) have a habitual or generic aspect,[71] whereas those of ($38b_i$) and ($38b_{ii}$) bear another aspectual nuance: that of progressiveness.[72]

On the other hand, the periphrastic conjugation V-PROSP + AUX is almost the only way to express the future tense, and although in some varieties the *-ke* forms can be used for this purpose, they carry an inherent sense of doubt. In any case, the relationship between the future and probability can also be seen in a more modern use of the periphrastic future ($39b_i$), a use which has been extended to the past tense ($39b_{ii}$). The following are some examples of this construction:

(39) a_i. *Gero* *hondartzara* *joan-en* *naiz*
later to.the.beach go-PROSP AUX.NPST
'I'm going to the beach later'

a_{ii}. *Ez* *dakit* *ekarri-ko* *du-en*
no I.know bring-PROSP AUX.NPST-CPL
'I don't know whether (s)he will bring it'

b_i. *Etxean* *egon-en* *da*
home-INES.SG be-PROSP AUX.NPST
'It may/will be at home'

---

71 Indeed, to express or strengthen the meaning of a "real" habit, the particle *ohi* is usually introduced. However, constructions without *ohi* have to be used in sentences such as *when I go...* (*...joaten naizenean*). In addition, there used to be an archaic way to express habitualness or gnomic aspect, using the verbs *joan* 'to go' (if intransitive) or *eroan* ~ *eraman* 'to carry' (if transitive) as auxiliaries (Irigoyen 1996): *puztu doa* 'it usually swells', *egin daroa* ~ *darama* '(s)he usually does'; cf. Sp. ***viene** a ser/siendo... **lleva** tiempo haciendo/siendo....* In addition, the verb *jakin* 'to know' is also used dialectally for that purpose: *badakie etorten* 'they usually come' (lit. 'they know coming') (Ortiz de Urbina *GB*: 296).

72 However, there is still a slight difference between the synthetic form and the periphrastic *+ari* one, which native speakers perceive and which prevents them from using the two forms interchangeably:

a. *Jendea dator* 'People are coming [at last]' (focusing on the factuality) ≠
b. *Jendea etortzen ari da* 'People are coming' (focusing on the ongoingness)

Another dialectal way to periphrastically express the progressive aspect is to add the *-ki* suffix to the verbal stem with the corresponding auxiliary: *joaki da* = *badoa* '(s)he is going', *ekarki du* = *badakar* '(s)he is bringing it', and so on.

b_{ii}.	*Hondartza-n*	*galdu-ko*	*genuen*
	beach-INES	lose-PROSP	AUX.PST
	'We may have lost it on the beach'		

9.7.1.3. Differences in morphology

The verbs which (also) work as auxiliaries are in fact, and unavoidably, the ones which show the highest frequency of use. This may be one of the reasons why they have grammaticalized in that way, and, at the same time, it is one of the consequences of this grammaticalization. All this has fostered the peculiar nature of these verbs with regards to their irregularities or archaic features in morphology (in the broadest sense of the word), as well as the diffusion/regularization of certain phonological changes to some other verbs. To clarify, here are some examples:

a. Irregularities in the root. The auxiliaries *izan*, and in some dialects **edun* (both also used as autonomous verbs) show more irregularities than most of the other verbs, probably due to the fact that frequency of use contributes to conservation against the usual analogical forces.[73] Thus, in the standard variety we have the following forms of *izan*:

(40)	*n-a-**iz*** 'I am'	*ba-n-in-**tz***, *ba-l-**itz*** 'if I ~ (s)he were'
	*h-a-**iza**-teke* 'thou may/will be'	*h-in-**tza**-teke*, *l-**itza**-teke* 'I ~ (s)he would be'
	*z-a-r-**a***, *d-ir-**a*** 'you ~ they are'	*g-in-**ø**-en* 'we were' (< **g-in-**a**-en*)[74]

To these forms, we must add the 3rd person singular of the present indicative (*da*), since it does not seem to contain the verbal root of *izan*, but appears to be simply the present tense prefix.

By contrast, the root of **edun* (commonly *-u-*) is irregular in some dialects, as a result of its merger with the present tense vowel. Thus, in Guipuscoan, when the object is 3rd p. sg. and the agent does not belong to the same person: *-e-* (*det* 'I have it', *dek* 'thou hast it', *dezu* 'you have it', *degu* 'we have it'). In Biscayan, there is an *-o-* in such forms and also in *nok* 'thou hast me' and *nozu* 'you have me'.

Finally, the auxiliary verb **edin* (once autonomous) also shows irregularities. While the root generally appears as *-di-* or *-i-* (*da-**di**-n* 'for (s)he

73 Yet there are local varieties, especially in the west, in which the analogy has entered these auxiliary verbs too, e.g. *nintzan* 'I was' : *gintzan* 'we were' ← *giñan*.

74 Cf. Old Biscayan *ginean* < **g-in-**a**-an*, with the western common dissimilation *a* + *a* > *ea*.

to...[INTR]' vs. *da-**i**-teke* '(s)he can...[INTR]'),[75] in the forms which have an indirect object (Nor-Nori) there is an apparent suppletion which comes from the contraction of the root plus the *dative flag* (**-di-ki-*): *da-**ki**-o-zu-n* 'for you to...[INTR] to him/her'.[76]

b. Irregularities in the tense marker (vowel). Even though this point is not exclusive to the auxiliary verbs, we can say that it mostly affects them. Some examples of this are related to those in the previous point, e.g. the *-o-*, *-e-*, *-u-* variants all come from **da-(d)u-* (cf. west. *dau* '(s)he has'), the second one specifically from an intermediate *-eu-* variant attested in some old texts (Laz. *deut* 'I have').

The paradigm of the verb *izan* has also dialectally suffered some similar changes, and the *-a-* of the present indicative for the 1st and 2nd pl. persons appears as *-e-* (*gera, zera* 'we are, you are', central dialect) or *-i-* (*gira*, *zira*, in eastern varieties).

Regarding the past forms of the verb **edun*, the fusion between the tense marker (*-en-*) and the pluralizer of the direct object (*-it-*) of the 1st or 2nd person is quite remarkable: *g-**int**-u-en* '(s)he had us', *z-**int**-u-da-n* 'I had you', from **g-**en**-**it**-(d)u-en* and **z-**en**-**it**-(d)u-en* respectively.

To conclude, the irregularity of the present tense vowel of the **ezan* auxiliary has yet to be solved. It is striking that even in the dialects in which the *-a-* has spread to the past tense, some conjugated forms of this verb also show an *-e-* in the present. Thus, for example, Leiçarraga wrote *d**e**zan* [3sgABS.3sgERG] and *n**e**zan* [1sgABS.3sgERG] (but *(h)**a**zan* [2sgABS.3sgERG]), Axular used *d**e**zan* but also *n**a**zan*, and Dechepare used *d**a**zan* at least once. In fact, it is hard to establish what the original state of affairs was.

c. Pluralizers. As has already been emphasized in the corresponding section, there is a pluralizer *-it-* whose origin remains unknown, and which is only used with the verbs **edun*, **ezan* and **iron*. On the other hand, another mysterious pluralizer, *-(i)r-*, only appears in some forms of the verb *izan* (*gara* 'we are', *zara* 'you are', *dira* 'they are').

d. Dative flag *-i-*. Although this allomorph can also be found in the dative or allocutive forms of other verbs, we can highlight its presence, either prefixed or suffixed (or even both at the same time, as in some varieties and in the standard one), with the tripersonal forms of the auxiliary verb **ezan* (cf. *d**i**e-*

75 Cf. the forms of *egon* such as *da-**go*** '(s)he is (at)' vs. *da-**u**-de* 'they are (at)' (< *da-**(*g)o**-de*).

76 In some old texts we find examples of *-kidi-*, in which the elements do not appear in the order we would expect, and which seems very likely to be the precedent form of the current *-ki-*, through an intermediate stage **-kii-*.

*za**i**on* ‘that (s)he do it to him/her’, *deza**i**on* ‘idem’, *d**i**ezon* (< **diezaon*) ‘idem’), and as a suffix in the bipersonal forms of *izan* (*za**i**o* ‘(s)he is to him/her’, *za**i**gu* ‘(s)he is to us’, *natza**i**zu* ‘I am to you’, etc.).

e. *-te* potential morpheme. This allomorph is only used with the verbs *izan* and **edin* (*date(ke)* ‘(s)he may/will be’, *zinate(ke)* ‘you would be’, *ninte(ke)* ‘I could...[INTR]’).

f. *z-* as a past tense marker. The morpheme *z-* appears as a past tense marker in central and eastern dialects, whenever the word-initial morpheme corresponds to the 3rd person. However, in the west *ø-* is found. This is one of the oddest features of the western dialect. Nevertheless, even in the west, *z-* appears in the forms of some verbs such as *irudi* ‘to seem’ or *jardun* ‘to be busy (with)’, but we should especially highlight the presence of this prefix in the past tense forms of *izan*, **edin* and **ezan*, as well as those of **edun* (especially with the plural object). These western past forms with *z-* are somewhat more abundant in old attestations.

9.7.2. Auxiliaries from synthetics, synthetics from Serial Verb Constructions (SVC)

As already mentioned, all auxiliary verbs of historical Basque are or were autonomous verbs with their own meaning (cf. Anderson 2006). Leaving aside the peculiarity of the realis/irrealis bipartition made by the verbs **edin* and **ezan*, we find that Basque does not differ greatly from certain other European languages such as French and Italian (modern Spanish, unlike Old Spanish, does not distinguish between intransitive and transitive auxiliaries). Cf. e.g.: Fr. *il **est** venue* ‘he has come’ with Basq. *etorri **da***, or Fr. *il **a** fait* ‘he has done’ with Basq. *egin **du***. English is also similar in that it uses the intransitive structure with any verb to render the progressive aspect: *(s)he **is** coming = etortzen ari **da***, or *(s)he **is** doing = egiten ari **da*** (vs. *egiten **du*** ‘(s)he (usually) does’).

As for the latter construction, Mounole (2011) has reconstructed a stage, probably prior to the emergence of *ari*, in which the progressive aspect would be expressed by means of the intransitive periphrasis **egi-te-n da* [do-NOMNL-INES COP] ‘(s)he is on doing’, equivalent to the one which yielded the modern present continuous in English (Trask 1995b: 226). The object of that *doing* would be expressed through an objective genitive: **ogi-aren egi-te-n da* ‘(s)he is on doing/making of bread’. Later, through analogy with the transitivity-sensitive perfective periphrasis (*joan **da*** ‘**(s)he has** gone’ vs. *egin **du*** ‘**(s)he has** done’), the progressive periphrasis would have acquired the same distinction, hence *joaten da* / **egiten da* → *egiten du* (with object in abs.: *ogi-a egiten du* ‘(s)he is mak-

ing bread'), which eventually became gnomic/habitual. The TOG phenomenon mentioned in §2.2.2.10 would support this hypothesis, inasmuch as the objective genitive still dialectally retained in other non-finite constructions such as *aita-ren* ***ikuste-ra*** *[noa]* '[I'm going] **to see** my father' (Trask 1995b: 225) would be a remnant of the former nominal nature of the lexical verb (*egi-te-*).

In turn, present-day synthetic verbs would come, according to Lakarra (2008a), from asymmetrical SVCs (Aikhenvald 2006: 22-28), where the so-called *minor verb* or AUX would have grammaticalized as a tense/aspect marker of the *major verb* or lexical verb (cf. prs./ipfv. *da-* < **dar* 'sit', pst./pfv. *-en* < **den* 'finish'), or as an argument introducer (such as the causative *-ra-* < **-la-* < **lo* < **don* 'take', cf. Lakarra 2012b, or the *dative flag -i-* < **nin* 'give'), or the same **dar* 'sit' as a kind of locative (now fossilized in non-finite forms such as *jauzi* < **e-da-dutz-i* 'to jump' vs. *utzi* < **e-dutz-i* 'to leave'). According to Aikhenvald (2006: 52), these constructions are a prominent feature of the languages of Southeast Asia, Oceania, New Guinea, and West Africa, although they can also be found in many languages across all the continents. Similarly, their use in most European-based creole languages is also noteworthy. These being so different from Basque, we can expect that, by means of holistic typology and comparative studies with SVC-languages, we will one day be closer to shedding light on some of the aspects concerning the origin and the evolution of the Basque finite verb morphology, which are so obscure today.[77]

9.8. Final remarks

In this final section, I shall try to sum up the most important aspects discussed until now. I shall also add some details in order to offer a panoramic view of the current state of the reconstruction of the Basque finite verb morphology.

9.8.1. Dialectal variants and reconstruction

On the one hand, it is true that both internal reconstruction and the reconstruction based on the comparison between dialects will not take us very far back in time, as the CB postulated by Michelena (1987 [1981a]) is usually dated to the 5th-6th centuries. Despite this, part of his reconstruction work (especially con-

77 As J. A. Lakarra points out to me, the aforementioned languages are often characterized by a rather poor flectional morphology, and a very limited amount of verbs with synthetic conjugation, if any at all. Both of these features can be assigned to PB, and the latter even to historical Basque.

cerning phonology, cf. *FHV*) had already reached previous dates, partly thanks to the Aquitanian onomastics, which date back to the first centuries of our era. But on the other hand, I consider it necessary to compare and to formally reconstruct the main dialectal variants, in order to establish the original forms, taking into account the analogical forces that must have played an important role both in the common system and in each dialect.

The first striking point in this issue is the difference between the paradigm of the auxiliary verb and those of the other verbs, in relation to their irregularities. This not only means a greater frequency of use of the auxiliaries, but it also assures us, in general, of their formal antiquity, and, in particular, of their morphological archaism.

Let us take the verb *izan* 'to be' as an example. In the indicative mood, its present tense forms are the most irregular ones, and they also vary considerably from one dialect to another, although, as we shall see, it is not hard to arrive at a common (yet also irregular) paradigm. Some historical forms are shown in Table 9.13.

Table 9.13. Present forms of the verb izan *'to be'*

	B	**G**	**HN**	**L**	**S**	
1sg	*nax*	*naiz*	*na(i)z*	*naiz*	*niz*	< *naiz* < **n-a-iza*
2sg	*ax*	*aiz*	*ya(i)z*	*haiz*	*hiz*	< *haiz* < **h-a-iza*
3sg	*da*	*da*	*da*	*da*	*da*	< *da* < **dar*
1pl	*gara*	*gera*	*gara*	*gare*	*gira*	< **g-a-ra*
2pl	*zara*	*zera*	*zara*	*zare*	*zira*	< **z-a-ra*
3pl	*dira*	*dira*	*dire*	*dire*	*dira*	< **d(a?)-i(-?)ra*

Curiously, out of all the forms shown, only those corresponding to the 1st and 2nd person singular clearly show the root of the verb *izan* (*-za-*), even though the final vowel is lost. On the other hand, the 3rd person singular form seems to be the grammaticalization as a copula of the same element from which the present/imperfective prefix *da-* developed, that is the proto-verb root **dar* 'sit' (Lakarra 2008a).

However, the plural forms contain an *-r-* which is hard to interpret. It has been suggested that it could be an instance of rhotacism (**-z-* > *-r-*), so that the *-ra* could belong to the *-za* verbal root,[78] but such a process would be unparalleled in Basque. In order to endow this reconstruction with some regularity, we

78 Nevertheless, we cannot dismiss the fact that we are dealing with another case of suppletion, where the plural forms would contain a different (unknown) root. Indeed, this would be the only instance of rhotacism in Basque, and hence it would not be a phonological rule (as it usually is when it occurs in a given language).

should expect that the *-i-* of the verbal stem which appears in *naiz* and *haiz* should also be found in the plural forms. However, there is no trace of such an *-i-* outside *dira*.[79]

As for the last form in the table, its first segment (*d-*) belongs to the present/imperfective *da-* morpheme, and the final *-ra* seems to be the same as the one we find in *gara* and *zara*. Therefore, the *-i-* could be the same as that of *naiz* and *haiz*, and we might reconstruct **daira* (< **da-i-za*?). This would mean, together with the lack of a clear plural marker in the forms *gara* and *zara*, that the 3rd pl. would not differ morphologically from the form we would expect for the singular one, which could show us a stage of the language in which number was not expressed (at least obligatorily) through a morphological element in the verbal complex.

In the case of **edun* 'to have', the reconstruction is quite simple in the present tense forms, but in the past ones we find more obstacles. In Tables 9.14.1 and 9.14.2, I put forward the forms that have a 3rd p. sg. (Table 9.14.1) or 3rd p. pl. object (Table 9.14.2), and the agent belonging to each person and number.

Table 9.14.1. Past forms of the verb *edun *'to have' with a 3rd p. sg. P*

	B	**G**	**HN**	**L**	**S**	
1sg	*neban*	*nuan*	*nuen*	*nuen*	*nian*	< **n-e-du-en/-an*
2sg	*euan*	*uan*	*uen*	*huen*	*hian*	< **h-e-du-en/-an*
3sg	*(z)eban*	*zuan*	*zuen*	*zuen*	*zian*	< **z-e-du-en/-an*
1pl	*genduan*	*genduan*	*gen(d)uen*	*ginuen*	*günian*	< **g-en-du-en/-an*
2pl	*zenduan*	*zenduan*	*zen(d)uen*	*zinuen*	*zünian*	< **z-en-du-en/-an*
3pl	*(z)eben*	*zuen*	*zuten*	*zuten*	*zien*	< **z-e-du-de-n*

Table 9.14.2. Past forms of the verb *edun *'to have' with a 3rd p. pl. P*

	B	**G**	**HN**	**L**	**S**	
1sg	*nebazan/nituan*	*nituan*	*nituen*	*nituen*	*nütian*	< **n-e-it-du-en/-an*
2sg	*euazan/ituan*	*ituan*	*ituen*	*hituen*	*hütian*	< **h-e-it-du-en/-an*
3sg	*ebazan/zituan*	*zituan*	*zituen*	*zituen*	*zütian*	< **z-e-it-du-en/-an*
1pl	*genduzan*	*genduzen*	*gin(d)uzen*	*ge/inituen*	*güntian*	< **g-en-it-du-en/-an*
2pl	*zenduzan*	*zenduzen*	*zin(d)uzen*	*ze/inituen*	*züntian*	< **z-en-it-du-en/-an*
3pl	*ebezan/zituen*	*zituen*	*zituzten*	*zituzten*	*zütien*	< **z-e-it-du-de-n*

79 The *-i-* of the variants *gira, zira* is usually explained through analogy with *dira*; the *-e* of *gare*, *zare* through reanalysis, from *garen* '(that) we should be', *garela* 'that we are', originally **gara-en*, **gara-ela*; and the *-e-* of *gera*, *zera* would come from a mixture of both, that is, **gara-en/la* > *gare(n/la)*, and through vowel methatesis > *gera*, probably favored by the form *dira*.

As we can see, the past tense suffix varies according to the dialect. At the westernmost and the easternmost ends, *-an* is used, whereas in the central varieties the variant found is *-en* (cf. §9.8.3). On the other hand, in the western dialects and, to a lesser extent, in the central ones, we can find variants (sometimes exclusively) with the pluralizer *-z(a)*, instead of *-it-*.[80] Moreover, even if historically the *-it-* forms are older and more numerous, in the westernmost varieties the use of the pluralizer *-za* in this verb and a number of others has become generalized. The antiquity of *-it-* is assured because it can also be found in some other related forms (of all dialects) in which the object does not belong to the 3rd person but to the 1st (*gintuen*) or the 2nd (*zintuen*) one (cf. Figure 9.5).

As far as the latter forms are concerned, Martínez-Areta (2012), in line with Oregi (1974), has proposed reconstructing them in exactly the same way as the ones shown above, in which the 1st and 2nd pl. are agents and the patient belongs to the 3rd pl. (*genituen*, *zenituen*). At first, the 1st and the 2nd person would have distinguished the number only through the formal difference of the agreement marker (sg. *n-*, *h-* vs. pl. *g-*, *z-*), but the 3rd person would not have grammaticalized this difference in the verb morphology (stage I). At a given time it would develop that distinction (stage II), and in the end, through analogy with the 3rd person, the forms with 1st or 2nd plural objects would also take the plural morpheme (stage III). This process is illustrated in Figure 9.5.

	Stage I	Stage II	Stage III
'We have it/them'	**g-en-du-en*	**g-en-(it-)du-en*	**g-en-(it-)du-en > gen(it)uen*
'(S)he has us'	**g-en-du-en*	**g-en-du-en*	**g-en-it-du-en > gintuen*

Figure 9.5. Possible development of *edun *past forms beginning with 1st p. pl. (A or P) agreement markers*

The difference in the development of the phonological form between the first and the second row can only be explained in terms of use frequency and/or analogy. While the form with the plural agent varies according to the number of the patient (*genuen* : *gen-it-uen*), the one containing that person (1st p. pl.) as a patient is always the same (provided that the agent is 3rd p. sg.), and hence the pluralizer *-it-* will always appear in it. Thus, by means of the expected phonological changes, the original form **genituen* '(s)he has us' would change to > **ge(h)ĩtuen* > **geintuen* > *gintuen*. The former homophone form *genituen* 'we have them' may have developed in the same way —in fact there are variants like *güntian* (< **güntüan* < **gintuan*) in Souletin and *giñun* (< *gindu(e)n* < *gintuen*)

80 To simplify, and since we have reasons to believe that *-it-* is older than *-za*, here I have used only the first one to represent the reconstructed forms.

in Guipuscoan, where the two forms have merged— but, through analogy with *genuen* 'we have it', it would have been reshaped as *gen-it-uen*.[81]

To conclude, as for the origin of the complex system of allocutive verbs, we can reconstruct three main sources which do not mutually exclude each other, but which are complementary. Firstly, we have the simple addition of the listener's agreement marker, either with the help of a morpho-phonological element, or without it (*dago* → *d(i)agok/zagok/jagok*). This procedure would come from the grammaticalization of a previous usage: the adding of the dative argument (objective at the beginning, but ethical later on). The new use kept the original morphology, while the common usage went a different way and innovated by adding *-ki* (instead of *i-*) to the common form (*dago*). As a result, non-allocutive forms such as *dagokik* appeared (but cf. western *jagok* in both senses).

Secondly, we have the root suppletion, in the origin of which we can find implicational mechanisms, but in this case they do not always belong to the *dative*: the verb *izan* 'to be' is generally replaced by a common form of the verb **edun* 'to have', with a 2nd person singular agent. Therefore, *da* 'it is' becomes *duk* 'thou hast it', and *naiz* 'I am' becomes *nauk* 'thou hast me'. However, in some dialects, certain forms of *izan* do not undergo this suppletion, and the agreement marker is simply added to them (*litzateke* 'it would be' → *litzatekek* 'id. (allocutive)'). But, in the same cases, in the varieties in which this process is regular for the whole paradigm, the resulting allocutive form does not belong to another common form of the verb 'to have'. For instance, *da* is to *duk* the same as *litzateke* 'it would be' is to *huke* 'thou would have it', but the latter (*huke*) is not the allocutive form belonging to *litzateke* in any dialect. Instead we have the aforementioned *litzatekek*, or the standard *lukek*, whereby the listener's agreement marker is simply added to the verb form *luke* '(s)he would have it'.

Finally, the verb **edun* has had two mechanisms to build its allocutive forms. These divide the current dialectal varieties into three big blocks, with the intermediate one being a kind of mixed solution. In the western varieties, the first type of derivation of allocutive forms has been generalized: the adding of the agreement marker with the help of the old *dative flag*: *dau* '(s)he has it' → **diauk* > *jok*, *nau* '(s)he has me' → *niauk*[82] > *niok/ñok/najok*. In the eastern ones, the prevailing solution was suppletion. The verb *-i-*, a three-person auxilia-

81 The form *ginituen* is also attested (for example in Axular) with both meanings; perhaps the frequency of the alternation betweeen *genuen* 'we have it' and *genituen* 'we have them', meant that the latter did not undergo the expected change (*-eni-* > *-ein-*) and, through analogy, the form **genituen* '(s)he has us' did not change, or even if it did, it would have been reshaped accordingly.

82 This form is not attested in the west, but it is in the earliest texts of the eastern dialects, although it has left no traces in the present-day dialect.

ry once common to all dialects, and related (according to Lakarra 2008a) to the verb *-i(n)-* 'to give', replaces **edun* in its allocutive forms: *du* '(s)he has it' → *dik* 'id. (allocutive)' (lit. '(s)he has it to thee'), *nau* '(s)he has me' → *naik*, etc. It is remarkable that only the forms with the 3rd person singular object of the *-i-* verb are used as non-allocutive auxiliaries (cf. the so-called *me lui* constraint). As a result, forms such as *naik* do exist, but only as allocutive forms. In the central varieties, by contrast, both systems appear mixed, the use of the suppletion being general when the object corresponds to the 3rd person, but varying whenever the object corresponds to the 1st or the 2nd person, also depending on whether it is singular or plural. Simplifying somewhat, we have the combinations shown in Table 9.15.

Table 9.15. Allocutive variants of *edun *verbs*

West	Central		East
Paradigm I	Paradigm II	Paradigm III	Paradigm IV
*naj**o**k* '(S)he has me (alloc.)'	*natxi**o**k*	*ni**o**k*	*nai**k***
*gaj**o**zak* '(S)he has us (alloc.)'	*gatxi**o**k*	*gait**i**k*	*gait**i**k*
*j**o**k* '(S)he has it (alloc.)'	*di**k***	*di**k***	*di**k***

9.8.2. Diathesis and agreement

In order to sum up what has been said in §9.5, and to outline a general picture of the different proposals and ideas about the diathetical distinctions of the Basque verb, we can say that there are two opposite tendencies. One defends the passivity of the conjugated forms (at first as a synchronic interpretation, and later, as a diachronic explanation of historical morphological facts), while the other shows skepticism or even rejection of this idea.

In the first group, there are classical authors such as Schuchardt (1972 [1893], 1947 [1923]) and Lafon (1999 [1954b]), as well as some other well-known linguists such as Rebuschi (1986) and Trask (1977). In the second group, we must include Luis Michelena and Joseba A. Lakarra (2005).

Recently, Mikel Martínez-Areta (2012) has rescued the position defended by Lafon and Trask, approaching it from a much more syntactic-typological perspective, in order to find out how the transitive forms were formed, and what could be the origin of such a diathetical distinction (intransitive vs. transitive). According to him, Proto-Basque would have been, at some point, similar to Chinese, in which the diathetical interpretation is made clear by the context, through the presence/absence of explicit arguments and/or certain passivizing elements. Then, by means of a generalization —whose type and motivations are

not yet clear— the passive constructions took on an active interpretation and, as a consequence, the current ergative-absolutive system emerged.

Even if this hypothesis is not entirely satisfactory, it can explain the data of the present tense conjugation, as well as some of the non-present tense ones. Among these, there is disagreement between the forms with a 3rd person patient and the rest of them, since, among the former ones the order of morphemes is not the same as in the present tense equivalents, and the agent agreement is prefixed in the same way as that of the intransitive subject. But there is a minor difference when the agent belongs to the 1st or the 2nd person singular. In these cases, there is no trace of the "mysterious" *-n-* (cf. *nuen* < ***n-e-du-en* 'I had it' vs. *ninduen* < ***n-en-du-en* '(s)he had me'), and it is in the plural forms that this element can be found again, so that the difference disappears in certain forms of certain dialects (cf. §9.8.1). Trask tried to find in this element a passivizing function or an intransitivity marker, but such an explanation does not clarify its distribution.

In view of these facts, Martínez-Areta draws the following conclusion: the alternation between *-n-* and *-ø-* would have had its origin in a phonological/allomorphic alternation. Thus, an *-n-* (belonging, according to him, to an oblique form of the pronouns) which in some contexts tends to disappear (there are certain phonological parallels), would be morphologized in order to mark the difference between those forms which originally had a passive interpretation and those with an active one.

As for the *split ergativity* (cf. §9.4.1.2), the author recalls —as did Ortiz de Urbina (1989)— that a system in which the present tense is ergative-absolutive and (part of) the past tense is nominative-accusative is actually at odds with typological data, which instead predicts that if a split is found, the past tense will be more ergative-absolutive than the present tense (cf. Dixon 1994). As an explanation for this, Martínez-Areta ventures to propose that, among the non-present tenses, the original meaning would be the so-called hypothetical, since it appears that the past tense forms are morphologically derived from it (***nu-ke*** 'I would have it', *ba*-***nindu*** 'if (s)he had me' → ***nu-en*** 'I had it', ***nindu-en*** '(s)he had me'). The author proposes a prospective aspect as the origin of the hypothetical, which, following Comrie (1981b), would be more *agent-oriented* than the imperfective one. In this way, in all of the imperfective aspect forms (→ present tense) the passive interpretation would be generalized, leading to the ergative-absolutive marking. By contrast, the forms with a 3rd person patient of that prospective aspect (→ past/irrealis) would keep their active interpretation, since this aspect is more *agent-oriented*. In the forms with 1st or 2nd person objects of the prospective, the passive interpretation would spread due to the animacy hierar-

chy, since the 1st and 2nd person pronouns are easier to topicalize, and therefore to interpret as patient subjects of passive constructions.

Finally, and after checking different languages for some phenomena related to the topic, such as information order and subject prominence, the author proposes assigning an SVO word-order to Proto-Basque, with a strong tendency to topicalization. As a result, we would actually have a TSV(O) order, from which the historical SOV would derive, by means of the aforementioned active reinterpretation of passive constructions (cf. §10.4.4).

An alternative account for the unusual split ergativity of Basque is given in Aldai (2000). In his proposal, Aldai reconstructs a Proto-Basque which was entirely ergative throughout its verb morphology (where, for example, the equivalent form of the historical —or reconstructed according to historical variations— ***n**e(*d)uen* would be **e(d)u**da**n*, with the same suffix marker of the ergative agreement that we see in present tense forms such as *du-**t*** < **da(d)u-**da***). Besides, as occurs in many languages of the ergative-absolutive type, this Proto-Basque would have had an antipassive device by which the agentive argument —usually encoded as ergative—, when P = 3rd p., would acquire the same expression of the common absolutive arguments, i.e., a prefix (which, following the example above, would be *n-*, just as it is historically: *n-uen*). Considering that an imperfective/perfective distinction would only be effective (and hence morphologized) in the past tense, Aldai justifies the fact that this antipassive has left a trace only in the non-present tenses by means of indirect parallels in other languages in which there is a certain association between imperfective and antipassive. This would have resulted in the historical split ergativity.

9.8.3. SVCs and a possible origin of the Basque synthetic verbs

Throughout this article, I have mentioned several times the so-called Serial Verb Constructions (SVC), whose presence at some point in the past of the Basque language is defended by the linguist Joseba A. Lakarra. Among the different types of SVC included in the typological bibliography (Aikhenvald & Dixon (eds.) 2006, Heine & Kuteva 2002), for Proto-Basque Lakarra has reconstructed various verb roots which were employed in asymmetrical SVCs: **dar* 'sit', **den* 'finish', **nin* 'give', **don* 'put, take (out)'.

These verbs (called *minor verbs*) would take part in constructions where they would be chained to others (called *major verbs*) without the mediation of any other morphological element of coordination or subordination, until they grammaticalized as affixes. In this way, **dar tor* [V V] would mean 'to sit-to come', that is, 'to be on coming'; the first verb was grammaticalized as *da-* with

the sense of progressive aspect (→ imperfective → present), and *da-tor* would have had the value of '(he/she/it) is coming', or, by adding a pronoun, **ni-da-tor* 'I am coming', **zu-da-tor* 'you are coming', etc. The pronouns were also cliticized, and they later merged with the tense marker, e.g. *na-tor*, **za-tor*.

Later on, the plural markers like *-za* were added to forms like the latter one, giving rise to **zator-za*, and from this, the modern *zatoz* (also attested as *zatorz*).

As for past and hypothetical forms, they both contain an *e(n)-* prefix which cannot yet be explained. We have two possible analyses: either it is a morpheme with two allomorphs (which have been used in a more or less complementary distribution), or they are two distinct morphemes (*e-* and *en-*) or the same one with and without an extension (*e-* and *e-n-*). The latter analysis does not only avoid suspiciously regular allomorphic variants, but it also unifies the analysis of one of the segments (*e-*). This *e-* can be related, as first suggested by Schuchardt, to the homophonous prefix of the verbal participles (*e-tor-ø/-i/-tzen*), although the *-n-*, which is inserted when the object does not belong to the 3rd person or when the subject belongs to the 1st or 2nd person plural, remains unexplained. In Table 9.16, the non-present stems of the verb **edun* from the standard variety are shown.

Table 9.16. Some junctures in past forms of the verb *edun*:* -n- *of the person marker + pluralizer* -it- + -d- *of the verbal root* *du(n)

3rd p. pl. P	**P other than 3rd p.**
-n(it)u- < **-n-e-(it-)du-*	-ni**n**du- < **-n-e(-?)n-du-*
-h(it)u- < **-h-e-(it-)du-*	-hi**n**du- < **-h-e(-?)n-du-*
-(it)u- < **-e-(it-)du-*	
-ge**n**(it)u- < **-g-e(-?)n-(it-)du-*	-gi**n**tu- < **-g-e(-?)n-it-du-*
-ze**n**(it)u- < **-z-e(-?)n-(it-)du-*	-zi**n**tu- < **-z-e(-?)n-it-du-*
-(it)u(z)te- < **-e-(it-)du(-z)-de-*[83]	

In the above mentioned paper, Martínez-Areta (2012) proposes reconstructing two groups of personal pronouns: the first group would be the one that has been merged with the present tense marker (*ni*, *hi*, *gu*, *zu* + *da-* > *na-*, *ha-*, *ga-*, *za-*) and the second one would have been grammaticalized for the non-present tenses (**nen*, **hen*, **gen*, **zen*). In this way, the original case division (nomina-

83 The *-z* of this form may appear to correspond to the homophonous pluralizer, but in this case it is actually an instance of "excresence" to avoid a *t-t* sequence (cf. *guzti* 'all', which according to Lakarra 2010b: 230 may come from *duzti* < **tuzti* < **tuti-* < Lat. *tōtiu(m)*; cf. also §5.11.2, *guzti* (19)).

tive/oblique)[84] would have been used for a tense/aspectual distinction. Later on, after the loss of the *-n* in some contexts, the variation *CVn-/CV-* would have become fixed and morphologized to mark the distinction between the absolutive (***nen****torren*) and the ergative (***ne****karren*) subjects.

Turning to the past tense suffix *-en*,[85] we now have Lakarra's proposal (2008a) according to which it would be a grammaticalization of another serial verb (**den*) whose meaning was 'finish', which would at first have marked the perfective aspect. The fact that it appears suffixed and not prefixed, as does the present tense morpheme, could be explained by iconic motivation in the order of the serial construction, that is, due to the fact that 'finishing' is always the last thing that happens. To put it in Aikhenvald's words (2006: 21, 35-36): "[t]he order of components in SVCs may match the temporal order of actions they denote", although she later adds that "[i]n SVCs which express grammatical meanings, for instance, aspect (…), the order of components follows grammatical rules rather than iconic principles".

By contrast, Gontzal Aldai (1998) proposes that the suffix which historically marks the past (or, in other words, which appears obligatorily in the forms of the past tense), must be related to the homophonous suffix of subordination (regarding this issue, in my opinion, it is important to keep in mind the almost symmetrical distribution of the variants *-en* and *-an*, which show both the past tense suffix and the subordination marker). Aldai takes as his starting point a remark made by Lafon on the conjugation of the past tense: it makes no difference between forms of indicative main clauses and of relative or subjunctive ones, with the marker never appearing repeated (e.g. past *zuen* : *zuen(**-*en)* vs. present *du* : *du-en*). In short, Aldai proposed that this suffix did not express the past, and nude forms such as **nu*, **zu*, etc. would be the ones which expressed that tense in prehistoric times. The frequent use of relativized forms and/or subordinated ones in narrative contexts (of past events) would thus have led to the reanalysis of that marker as the morphological expression of the past tense, but without losing its original meaning, clearly differentiated in present tense forms.

Another serial verb reconstructed by Lakarra (2008a) is **nin* 'give', whose own conjugation —with its own meaning— is attested especially in old texts, although some forms have survived until now in fossilized phrases (greetings, prayers, popular songs). But, until very recently, that root had not been recognized (in its most reduced form *-i-*) in any other part of the grammar, such as the ditransitive auxiliary of the indicative of the central dialects, or the dative case

84 Indeed, Martínez-Areta suggests explaining the irregular free pronoun form *ene* [1sg.GEN] as < **(n)en-e* [1sg.OBL-GEN] (see also §7.2.3).

85 But cf. the *-an* variant, at both ends of the Basque territory.

marker *-(r)i* of the nominal inflection. These grammaticalization processes are not at all extraordinary, if we consider the corresponding typological data.

As I already pointed out in §9.5.4, it is also possible that the so-called *dative flag* (at least in its prefixed variant *i-*; cf. *-ki(-)*, always as a suffix) could be a development of that verb **nin* 'to give', although the distribution between *i-* and *-ki* (< *-gi*) still remains unexplained.[86] On the other hand, in the above mentioned paper (Lakarra 2008a: 482) a relationship between the verb **nin* 'give' and the perfect participle *-i* suffix is suggested, but we have not found any evident typological parallel (cf. Heine & Kuteva 2002).

To conclude the discussion, in an unpublished paper Lakarra (2012b) has recently proposed a new interpretation of the synthetic causative *-ra-*.[87] The formal coincidence of its previous state **la-* with the expected phonological development of the locative morpheme (→ imperfective) *da-* in the absolute initial position would only be an illusion. This was due to a morpho-phonological change which affected —during an indefinite period of the Middle Ages— the last vowel of the first element of a derivation or compound: *-o/-e* > *-a*, cf. *luze* 'long' → *luza-tu* 'extend', *baso* 'forest' → *basa-ti* 'wild'. That is, its oldest form would not be **la-*, but **le-* or **lo-*. According to Lakarra, **lo-* could be derived from the root **don*, with a meaning that might resemble 'take', a verb that can grammaticalize as a causative (Heine & Kuteva 2002: 286).

Therefore, we can say that Basque was once an isolating language with SVCs whose minor verbs cliticized and grammaticalized as auxiliary verbs, and which began to become more agglutinative, with tense/aspect and causative-transitivizing morphemes. The mechanism of creating synthetic verbs fossilized at some point, and only the most frequently used ones survived. At the same time, some of those surviving synthetic verbs were grammaticalized as auxiliaries in new periphrases. This has turned into the productive pattern, which has become increasingly more widespread since the earliest written records.

---

86 Lakarra (2008a: 482, fn. 77) seems to suggest a genetic relationship between *i-* and *-ki* (< *-gi*). Perhaps we should understand the voiced velar consonant as an epenthesis in an intervocalic context: *dagogigu* 'it concerns us' (from Uriarte) < **dago.igu* < **da-go-(h)i-gu* < **da-go-ni(n)-gu*. The most common variant with a voiceless velar (*dagokigu*) would come from the reanalysis of the same form with a plural object (*dagoʒkigu*), in which the consonant has been unvoiced because of the preceding sibilant consonant.

87 Prior to this paper, he believed that *-ra-* 'causative' [< '*applicative'] < **-Vla-* ← **#la-* < **#da-* < **dar* 'sit', like the 'locative' *da-*, which had been grammaticalized also as an imperfective marker (> present). Now, he rejects that problematic double grammaticalization of the serial verb **dar* 'sit'.

10. Word order

Urtzi Reguero-Ugarte
(University of the Basque Country, UPV/EHU)

10.1. Introduction*

In this chapter, the order of constituents in the Basque clause, as well as within NPs and adpositional phrases, will be described and discussed. Firstly, I shall discuss the position of Basque within the parameter of attested constituent order patterns (SOV, SVO, VSO, etc.), and the relative positions of V and O (but also of other complements of the verb) (Section §10.2). Secondly, the word orders within the NP such as AdjN / NAdj, GN / NG, RelN / NRel, NDet / DetN, and within the adpositional phrase, i.e. NPo / PrN, will be examined (Section §10.3). The analysis will mainly focus on diachronic research. However, as will be shown, in both the constituent order at clause level and in some orders within the phrase, dialects do not show easily perceptible variations, neither between themselves nor in the course of their attested history.

Word order has played an important role in the linguistic typology of the last few decades. However, if we leave aside the generative literature, studies on Basque word order from a functional-typological perspective have been relatively scarce. On the synchronic level, the most systematic research is probably Osa-Unamuno (1990), in which word order in Basque is studied according to its communicative function, and hence from a functionalist perspective. On the diachronic level, due to the relative similarity of word order features across Basque dialects and over the attested history, comparison cannot provide tools to reach conclusions, and Bascologists have had to resort to internal reconstruction to try to get a glimpse of Basque word order in

* This work has been carried out thanks to a Grant for the Training of Researchers awarded by the University of the Basque Country / Euskal Herriko Unibertsitatea (UPV/EHU), with the code PIF10/2010/PIF10017. This publication is part of the Consolidated Group "Historia de la Lengua Vasca y Lingüística Histórico-Comparada" (HLMV-LHC), Ref. GIC 10/83, IT 486-10, and is also part of the Project "Hizkuntzalaritza Teorikoa eta Diakronikoa: Gramatika Unibertsala, Hizkuntza Indoeuroparrak eta Euskara (HiTeDi) / Lingüística Teórica y Diacrónica: Gramática Universal, lenguas indoeuropeas y lengua vasca (LingTeDi)", of the University of the Basque Country / Euskal Herriko Unibertsitatea (UPV/EHU), coordinated by Myriam Uribe-Etxebarria, with Ref. UFI11/14. I would also like to thank Mikel Martínez-Areta and Joseba Lakarra for their helpful comments, and to Mikel Martínez-Areta for the translation of the text.

prehistorical and precommon phases. Two attempts in this direction are Trask (1977, *HB*), and Gómez (1994) and Gómez & Sainz (1995). In addition, by adopting a holistic approach, Lakarra (2005, and later works) has proposed a complete reversal of both clause-level and intra-phrasal parameters from the VO type to the OV type. Finally, Martínez-Areta (2011, 2012) has rescued Trask's view, and added new elements to it. All these views will be discussed in section §10.4.

The purpose of this chapter, then, is to simply recapitulate the main theories on Basque word order, not to propose a new one.

10.2. OV vs. VO. Sentence constituent order in Basque

Basque is usually considered to be an SOV language. This order is not rigid, however, and some analysts have proposed that the most natural order in Basque is not SOV but SVO (see §10.2.3). Since the S is indisputably —at least in unmarked constructions— at the beginning, the first issue to be tackled is when we find OV, and when VO. The conditioning factors which trigger either order can be classified into two groups: 1) purely syntactic conditioning; and 2) pragmatic conditioning. Let us look at each of them in turn.

10.2.1. Syntactically conditioned constituent order

If the discourse is conditioned by the context, then there may be some reason for the speaker to stress a given constituent of a particular sentence. This can spark a marked, context-conditioned order, which will be commented on in §10.2.2. But if no constituent is to be stressed, then the unmarked order is typically OV in Basque affirmative sentences with a light O. This OV order, however, is not rigid. Purely syntactic factors, such as a heavy O, can favor a VO order. But there is no rule which makes this change predictable, and VO can actually be found in contexts which are neutral from both a syntactic and a pragmatic point of view, for reasons which are not easy to determine.

Let us consider several possible syntactic contexts, in order to study how the OV / VO parameter behaves in each one.

10.2.1.1. Affirmative sentences

In syntactically and pragmatically neutral contexts, Basque is an SOV language. Along with ergativity, this is one of the most outstanding features of this

language in comparison with the rest of the languages of western Europe. It should be remembered that the Basque verb can be synthetic (a single finite verbal form gathers the lexical meaning and the TAM markers, as well as the argument indexes), or periphrastic. In the most typical periphrastic construction (to which I shall limit my analysis), a finite form gathers some TAM features and the argument indexes, while a non-finite form gathers some tense/aspect features and the lexical meaning (see §8.2, §9.3, §9.4, §9.7). In both synthetic and periphrastic constructions of this kind (in which the historical order is non-finite form + finite form), the unmarked order of constituents in affirmative sentences with a standard (not too heavy) O is SOV. Since Basque is a subject pro-drop language, we often find simply OV. Here are two examples:

(1) a. *Bihar* *klase-a* *dauka-t*
tomorrow class-ART have-1SG
'I have a class tomorrow'

b. *Liburu-a* *irakurr-i* *du-t*
book-ART read-PFV AUX-1SG
'I have read the/a book'

Apart from the O (Direct Object), other types of verbal complements also present unmarked (S)CompV order in the same syntactic context, such as adverbial complements (2a, b, c), subject complements (3a, b), etc.[1]

(2) a. *Amona-ren* *etxe-ra* *n-oa*
grandmother-GEN.SG house-ALL.SG 1SG-go
'I go to grandmother's house'

b. *Haize-a* *hegoalde-tik* *dator*
wind-ART south-ABL.SG comes
'The wind comes from the south'

c. *Ni* *gaur* *goiz-ean* *merkatu-ra* *joan-go* *naiz*
I today morning-INES.SG market-ALL.SG go-PROSP AUX
'Today in the morning I am going to go to the market'

1 In fact, all kinds of complements behave in the same way as Os —themselves complements— do with respect to word order, although only occasional reference will be made to them in the rest of this paper.

(3) a. *Patxi* *altu-a* *da*
p.n. tall-ART is
'Patxi is tall'

b. *Oso* *nekatu-ta* *na-go*
very be.tired-PTCP 1SG-be
'I am very tired'

However, even in such contexts, OV or CompV is not obligatory. Provided that the finite verb does not begin the sentence, all of the examples in (1-3) would also be grammatical with a VO or VComp order. The variation between the two possible orders does not entail any predictable and easily definable value, not even in pragmatic terms.

This is the case in all of the dialects from the beginning of the tradition, as the following sample sentences show:

(4) a. *Guré* *Ienko* *hon-ak (...)* *mundu* *haur* *goberna-tzen* *du*
our God good-ERG.SG world this govern-IPFV AUX
'Our good God governs this world' (Tt., *OHB* 157)

a'. *Bizi-ak* *eman-en* *dro* *meritu* (Tt., *OHB* 160)
life-ERG.SG give-PROSP AUX merit
'Life will give him merit'

b. *mai-a* *bedeinka-tu* *begi* (Cap., *DC*)
table-ART bless-PFV AUX
'Let him bless the table'

b'. *ofrezi-etan* *deu-su-da-z* *neure* *obra* *guzti-ak*
offer-IPFV AUX-2SG-1SG-3PL my deed all-ART.PL
'I offer you all my deeds' (Cap., *DC*)

c. *Hauti-a* *guré* *da* (Tt., *OHB* 70)
choice-ART our is
'The choice is ours'

c'. *Hura* *da-té* *adeskide* *hon-a* (Tt., *OHB* 129)
he be-PROSP friend good-ART
'He will be a good friend'

d. *beste* *zazpir-ak* *gorpuz-eko-ak* *dira* (Cap., *DC*)
other seven-ART.PL body-L.GEN.SG-ART.PL are
'The other seven (= commandments) are the ones of the body'

d'. *leenengo* *zazpir-ak* *dira* *espiritual-ak* (Cap., *DC*)
first seven-ART.PL are spiritual-ART.PL
'The first seven (= commandments) are the spiritual ones'

This variation always occurs between the V and the element with the most immediate dependence on it. If, for instance, a transitive sentence has no O because this is contextually implied —recall that Basque does not have the possibility of marking these by means of anaphoric clitics— or for any other reason, the above variation can be between the V and the indirect object, as in the following examples:

(5) a. *Baina* *emazte-ari* *erran* *ziez-o-n* (Lç., *Luk* VII, 50)
but woman-DAT.SG say AUX-3SG-PST
'But he said to the woman'

a'. *Halaber* *erran* *ziez-e-n* *diszipulu-ei* (Lç., *Luk* XVII, 22)
likewise say AUX-3PL-PST disciple-DAT.PL
'Likewise, he said to the disciples'

b. *Silbia-k* *bere* *ugazaba-ri* *onela* *esan* *deus-o* (Laz., <1147r>)
p.n.-ERG his lord-DAT.SG so say AUX-3SG
'Silvia has spoken in this way to his Lord'

b'. *esan* *eus-a-n* *Silbero-ri* (Laz., <1153r>)
say AUX-3SG-PST p.n.-DAT
'He said to Silvero'

Thus, in pragmatically unmarked contexts, the axis of the sentence is constituted by the V and its complement (whether O or of any other type) or in any case the element syntactically closest to it, and the relative order between these

two elements. Adjuncts are located at the periphery of that bicephalous nucleus; if before, as an ambience for the main information, or if after, to add information or as an afterthought.

10.2.1.2. Negative sentences

Negative sentences in Basque have an order of elements which differs from the neutral one found in affirmative sentences. If the verb is synthetic, the standard order is *ez* 'not' + V + O (see (6a)). If the verb is periphrastic, the standard order is *ez* + V_F + O + V_{NF} (see (6b)):

(6) a. *Ez* *dauka-t* *diru-rik*
not have-1SG money-PTV
'I have no money'

b. *Ez* *du-t* *diru-a* *ikus-i*
not AUX-1SG money-ART see-PFV
'I have not seen the money'

In (6b), however, the order *ez* + V_F + V_{NF} + O (*Ez du-t ikus-i diru-a*) would not be ungrammatical either, just as in the examples (1a, b) above the order VO would also be possible.

From the earliest texts, however, another word order is also found with periphrastic verbs, which appears to be older. This order is as follows: V_{NF} + *ez* + V_F, the O being before or after the verbal complex. Here are some examples:

(7) a. *ni* *erruki* *ez* *na-u-zu* (Laz., <1144r>)
me take.pity.on not 1SG-AUX-2SG
'You don't take pity on me'

b. *gorde* *ez* *z-it-u-e-z-an* *aren* *Mandamentu* *santu-ak*
keep not ~ AUX his commandment holy-ART.PL
'They did not keep his commandments' (Cap., *DC*)

c. *Onerizte* *gabe-z* *uz-ten* *e-zitu-t* (Dch. X, 12)
love lack-INSTR leave-IPFV not-AUX-1SG
'I don't abandon you for lack of love'

d. *gu-ek* *ajutu* *ez* *du-gu* (Em. Gloss, ca. 1000 AD)
we-ERG ? not AUX-1PL
'We do not cast ourselves (?) / We do not have it as suitable (?)'[2]

However, the new order, with the V_{NF} after the V_F, and usually also after the O, also appears from the earliest texts, as in (8a, b):

(8) a. *halako-er* *so* *hon* *bat* *e-niro* *egin-en*
like.that-DAT.PL look good a not-AUX do-PROSP
'At ones like that I would not have a look' (Tt., *OHB* 112)

b. *alan* *ez-to-zu* *pekatu-rik* *egin-go* (Cap., *DC*)
so not-AUX-2SG sin-PTV do-PROSP
'That way, you won't sin'

The two orders, then, seem to be competing in the earliest texts, although eventually the *ez* + V_F + V_{NF} type has prevailed. Despite this, the old order V_{NF} + *ez* + V_F is still found as a fossil in some fixed expressions (as in (9a)), or in negative embedded clauses (as in (9b)). In these, since V_F is at the right edge of the clause, the relative suffix *-n*, the complementizer *-la*, or any other subordinating item can be attached to it with no need for the embedded clause to be cleft into two parts, as would obligatorily be the case with the order *ez* + V_F + V_{NF} (as in (9b')):

(9) a. *Ezagutu-ko* *ez* *n-uen* *ba!*
know-PROSP not 1SG-AUX indeed
'How shouldn't I know him?'

b. *[Ikus-ten* *ez* *du-zu-n]* *gizon-a*
see-IPFV not AUX-2SG-RLV man-ART
'The man that you don't see'

b'. ***[Ez* *du-zu-n]* *gizon-a* *[ikus-ten]*

2 The translation depends on the meaning that we assign to *ajutu*, the reading and interpretation of which is controversial (cf. *TAV*: 41-44). In the second interpretation, *ajutu* ought perhaps to be considered an object complement rather than a non-finite verb, but in transitive verbs, perfective non-finite verbs actually come from object complements, so that *guek ajutu ez dugu* would be an earlier phase of the structure V_{NF} + *ez* + V_F, and hence an early attestation of that order anyway.

10.2.1.3. Sentences with a heavy O

As I have said, Basque tends towards OV, especially when the O is light and the context is not pragmatically conditioned. However, in these pragmatically non-conditioned contexts, VO can also be found. This appearance is not easily predictable or definable, but, *ceteris paribus*, the longer the O is, the greater the probability of finding VO instead of OV will be. In other words, if the O is heavy, it will tend to move to the postverbal position. This can occur, for instance, if the O is a completive clause. Consider (10a, a', b, b'):

(10) a. *Patxi-k* *txiste* *bat* *konta-tu* *di-t*
p.n.-ERG joke a tell-PFV AUX-1SG
'Patxi has told me a joke'

a'. *Patxi-k* *konta-tu* *di-t* *txiste* *bat*

b. *Patxi-k* *[atzo* *bere* *lagun-ekin* *hondartza-ra* *joan*
p.n.-ERG yesterday his friend-SOC.PL beach-ALL.SG go

ze-la] *konta-tu* *di-t*
AUX-CPV tell-PFV AUX-1SG
'Patxi has told me that yesterday he went to the beach with his friends'

b'. *Patxi-k konta-tu di-t [atzo bere lagun-ekin hondartza-ra joan ze-la]*

(10a', b') are the VO versions of (10a, b). As a matter of fact, the four sentences of (10) are grammatical, but while (10a) sounds more natural than (10a'), (10b') sounds somewhat more natural than (10b). The reason seems to be that in (10a, a') the O is just an NP (of the structure 'noun + determiner'), whereas in (10b, b') the O is a whole completive clause. Let us now consider (11a, b):

(11) a. *Jon-ek* *ba-daki* *[Miren-ek* *esan* *di-o-la*
p.n.-ERG COND-knows p.n-ERG say AUX-3SG-CPL

[azterketa-rako *ikasi-ko* *du-ela]]*
exam-DST.SG study-PROSP AUX-CPL
'Jon knows that Miren has told him that she will study for the exam'

b. *?? Jon-ek [Miren-ek [azterketa-rako ikasiko du-ela] esan di-o-la] ba-daki*

(11b) would be the correct option if Basque were a language like Japanese, with a strict SOV order. But, in Basque, although sentences like (11b) are not ungrammatical, they are never used, for they would distort both the hearer's comprehension and the speaker's discourse flow.

The tendency to avoid heavy Os, or other kinds of Comp.s, to the left of the V can also be observed with Os other than completive clauses, as for example with relative clauses, or Os / Comp.s which are heavy for whatever reason. This is the case from the earliest written texts. In (12), we have some examples of heavy Os / Comp.s to the right of the V. In (12a), the postverbal element is a completive clause, in (12b) a relative clause, and in (12c) quite a long subject complement:

(12) a. *Franzia oro-k aithor-tzen du, [zu iaun handi*
France all-ERG confess-IPFV AUX you (sg.) lord great

bat, Baroin handi bat, Markis handi bat
a baron great a marquess great a

zire-la] (Tt., *OHB* 2)
are-CPL
'The whole of France confesses that you are a great lord, a great baron, a great marquess'

b. *[egun a-tan gogo-ra etorr-i*
day that-INES.SG mind-ALL.SG come-PFV

ia-ka-z-(a)n, esan ditu-z-(a)n, edo egin
AUX-3SG-3PL-RLV say AUX-3PL-RLV or do

ditu-z-(a)n] gauza guzti-ak ekarr-i begi-z
AUX-3PL-RLV thing all-ABS.PL bring-PFV AUX-3PL

memoria-ra (Cap., *DC*)
memory-ALL.SG
'Let him recall (lit. bring to the memory) all things which have come to his mind on that day, and which he has said or done'

c. *Han-go* *Bizkonde-ak* *eta* *seme* *guzti-ak*
there-L.GEN viscount-ABS.PL and son all-ABS.PL

iza-tu *dira* *bethiere,(...)* *errege-z* *enplegatu-ak,*
be-PFV AUX always king-INSTR.INDEF use-ABS.PL

estimatu-ak *fin* *eta* *leial* *frogatu-ak* (Ax., 3)
esteem-ABS.PL finely and loyally proven-ABS.PL
'All viscounts and their sons from there have been used, esteemed, finely and loyally proven by the king'

However, this is not a rule, and heavy Os or Comp.s can also be found to the left of the V. In this respect, the state of affairs hardly differs from the beginning of the tradition to the present day. (13a, b) are examples of heavy Os (completive clauses) in 16th-17th century texts, whereas (13c) is an example of a (relatively) heavy O in a recent novel by Bernardo Atxaga:

(13) a. *[buru-an* *min* *zue-la]* *erran* *zu-en* (Ax., 33)
head-INES.SG pain AUX-CPL say AUX-PST
'He said that he had a headache'

b. *zerren* *[milla* *bider* *geiago* *da-la* *da-n* *baño]*
because thousand time more is-CPL is-RLV than

esan-go *dab-ee* (Zumarraga's letter; cf. *ContTAV*: 97-102)
say-PROSP AUX-3PL
'(...) for they will say that it is a thousand times more than it is'

c. *Philipe Lou* *izen-(e)ko* *boxeolari* *bat-en* *hamaika*
p.n. name-L.GEN.SG boxer a-GEN.SG eleven

xehetasun *eman* *zi-zki-da-n* (B. Atxaga's novel, 2003)
detail give AUX-PL-1SG-PST
'He gave me countless details about a boxer called Philipe Lou'

According to Martínez-Areta (2011), the reason for avoiding Os which are too heavy is related to the economy of discourse processing. A number of experiments (cf. Hawkins 1983: 102-103, Grosu & Thompson 1977, Manning 1996) have shown that 'heavy O + V' structures can distort comprehension of the sen-

tence, because the hearer has to process a part of the predicate which is too large, without knowing what its nucleus, the V, is. As a result, many OV languages —or even VO languages with OV structures of some kind— have developed strategies to avoid them, such as allowing VO if the O is heavy (as e.g. with completive clauses) or splitting the O by means of extraposition, among others. The Basque phenomena, then, are perfectly compatible with cross-linguistically attested data.

10.2.1.4. Subordinate clauses

Here I comment on a final issue regarding syntactically conditioned phenomena: the order within embedded clauses. These can be of various types in Basque, but most are formed by the addition of a subordinating suffix to the finite verb of the embedded clause. The embedded clause can be, *inter alia*, completive by means of the suffix *-la* (14a), causal by means of the subordinating suffix *-lako(z)* (14b), or temporal by means of the subordinating suffix *-nean* (14c):[3]

(14) a. *[Nekatu-ta dago-ela] esan di-t*
tired-PTCP is-CPL say AUX-1SG
'(S)he has said to me that (s)he is tired'

b. *Nekatu-ta na-go, [korrika egin du-da-lako]*
tired-PTCP 1SG-be running do AUX-1SG-CS.SB
'I am tired, because I have been running'

c. *[Nekatu-ta na-go-(e)nean,] ohe-ra n-oa*
tired-PTCP 1SG-be-TMP.SB bed-ALL.SG 1SG-go
'When I am tired, I go to bed'

Since the subordinating suffix has to be attached to the finite verb of the embedded clause, but is somehow a linkage to the main clause, we would expect embedded clauses to favor OV order, in such a way that the V_F-Sb.Suff. complex has a syntactically demarcative function. Otherwise, the embedded clause would have to be split, and the subordinating suffix would no longer be in contact with the main clause. In fact, in some cases, such as when a relative clause

3 The temporal *-nean* is the relativizer *-n* plus the inessive sg. *-ean* in a temporal sense. In fact, all subordinating suffixes attached to finite verbs are derivations of the relativizer *-n* or of the completive *-la*. The purely relative embedded clauses will be commented on in §10.3.4.

modifies a head noun, the V_F-Sb.Suff. complex must be at the end of the embedded clause, as shown by the ungrammaticality of (15b):

(15) a. *[Ipuin-a irakurr-i du-en] ume-a ikus-i du-t*
tale-ART read-PFV AUX-RLV child-ART see-PFV AUX-1SG
'I have seen the child who has read the tale'

b. ***[Irakurr-i du-en]ume-a [ipuin-a] ikus-i du-t*

Note that in a main clause, the order VO, as in *Ume-ak irakurr-i du ipuin-a* 'The child has read the tale', would be perfectly grammatical, although the unmarked one is, in principle, OV.

However, other embedded clauses do allow Os, any other kind of complement, and adjuncts, to the right of the V_F-Sb.Suff. complex, as in the following temporal clauses:

(16) a. *[Ikus-i zu-enean] [txakur bat,] ikara-tu zen*
see-PFV AUX-TMP.SB dog a scare-PFV AUX
'When (s)he saw a dog, (s)he was scared'

b. *[Ikus-i zu-enean] [txakur bat zetorre-la] ikara-tu zen*
see-PFV AUX-TMP.SB dog a came-CPL scare-PFV AUX
'When (s)he saw that a dog came, (s)he was scared'

In both (16a) and (16b), but especially in (16a), it would be somewhat more natural to move the O —which in (16b) is a completive clause— to the left of the V_F-Sb.Suff. complex, so that this delimits the end of the embedded clause. However, (16a, b) are perfectly grammatical. In such cases, then, the OV ~ VO variation functions just as it does in main clauses. Thus, whether the OV order, whereby the embedded clause ends with its finite verb, facilitates the economy of discourse or not, can be analyzed in quantitative terms alone (e.g. by comparing the ratios OV / VO in embedded and main clauses in the speech of a given author, or of a given work or text, etc.).

10.2.2. Pragmatically conditioned constituent order

The arguments and examples set out in §10.2.1 were context-neutral. However, the order of constituents can depend on the context and the element which the speaker wants to stress. The basic rule is as follows: the focalized constituent or rheme, which in Basque grammar is called the *galdegaia*, is usually located im-

mediately before the V (whether simple V_F or periphrastic V_{NF} + V_F). That element can be any constituent of the sentence, whether it be the subject, the object (direct or indirect), any other complement, an adjunct, or even the verb itself.

A given constituent is the rheme or *galdegaia* of a sentence when it corresponds to the interrogative phrase of a question. In (17a, b), I present two sentences, first in a context-neutral order, and then as an answer to a question:

(17) a_i. *Ni-k ogi-a eros-i du-t* (Neutral)
I-ERG bread-ART buy-PFV AUX-1SG
'I have bought the bread'

a_{ii}. *Nor-k eros-i du ogi-a?* (Question A)
who-ERG buy-PFV AUX bread-ART
'Who has bought the bread?'

a_{iii}. *NIK erosi dut ogia* (Answer A)
'It is me who has bought the bread'

a_{iv}. *Zer eros-i du-zu zu-k?* (Question B)
what buy-PFV AUX-2SG you-ERG
'What have you bought?'

a_v. *OGIA erosi dut nik* (Answer B)
'It is the bread that I have bought'

a_{vi}. *Noiz eros-i du-zu ogi-a?* (Question C)
when buy-PFV AUX-2SG bread-ART
'When have you bought the bread?'

a_{vii}. *GOIZEAN erosi dut ogia* (Answer C)
'It is this morning that I have bought the bread'

b_i. *Jon kale-an dago* (Neutral)
p.n. street-INES.SG is
'Jon is on the street'

b_{ii}. *Nor dago kale-an?* (Question A)
'Who is on the street?'

b_{iii}. *JON dago kalean* (Answer A)
'It is Jon who is on the street'

b_{iv}. *Non* *dago* *Jon?* (Question B)
where is p.n.
'Where is Jon?'

b_{v}. *KALEAN dago Jon* (Answer B)
'It is on the street where Jon is'

The elements in capital letters are also prosodically stressed. Note that in ($17a_{iii}$, b_{iii}) the rheme is the subject, in ($17a_{v}$) it is the O, in ($17b_{v}$) it is the complement, and in ($17a_{vii}$) it is an adjunct.

However, just as the SOV order is not strict in context-neutral sentences, neither is the rheme-V order strict in pragmatically conditioned sentences. As answers to ($17a_{ii}$, a_{iv}, b_{ii}, b_{iv}), (18a, b, c, d) would be equally valid, provided that the rheme is prosodically stressed:

(18) a. *Ogia erosi dut NIK*
b. *Nik erosi dut OGIA*
c. *Kalean dago JON*
d. *Jon dago KALEAN*

In any case, the rheme cannot be located in any position of the clause. It is not allowed, for instance, to the left of the preverbal constituent. Thus, as an answer to ($17a_{ii}$), ***NIK ogia erosi dut* would be ungrammatical, etc. According to Ortiz de Urbina (2002), if the rheme is not in the preverbal position, then it must be at the end of the sentence, where the intonation falls.

These phenomena are by and large common to all dialects from the beginning of the tradition, although some regions have developed particular features.[4] Continental dialects, for example, have a strategy whereby if the V

4 An interesting issue is what happens in Wh-questions. In these, the pragmatically most relevant constituent is, by definition, the wh-constituent. Hence, we should expect Wh-questions in Basque to most typically have a #Wh-V- order, as is the case in ($17a_{ii}$, a_{iv}, a_{vi}, b_{ii}, b_{iv}). This is, in fact, the tendency which has increasingly generalized over the attested history, and is the grammatical one in Standard Basque. However, old texts of every dialect abundantly attest interrogative sentences such as Lç. *Zein-ek bekhatu egin du?* 'Who has sinned?' (lit. 'Which one has done a sin?'), where the order is actually Wh-O-V. The issue of how the adjacency Wh-V- in Wh-question has developed and

following the rheme is periphrastic, the usual V_{NF}-V_F order is changed to V_F-V_{NF}. Thus, the answer to ($17a_{ii}$) would not be ($17a_{iii}$), but *NIK dut erosi (ogia)*. On the other hand, peninsular western dialects have the possibility of turning the verb into the rheme by reinforcing it with *egin* 'to do'. Thus, if instead of ($17a_{ii}$, a_{iv}, a_{vi}) we asked *What have you done with the bread?*, the correct answer would be *(Nik) (ogia) EROSI egin dot (B) / det (G)* 'It is buying what I have done with the bread'.

10.2.3. Theoretical views on the sentence constituent order in Basque

The two basic rules of sentence constituent order in Basque, then, can be summarized as follows: 1) in context-neutral sentences, the unmarked order is usually SOV; 2) in pragmatically conditioned sentences, the unmarked order is usually rheme-V. As the word *usually* denotes, one should speak about tendencies rather than rules. In fact, it cannot be predicted whether in a given context-neutral sentence we will find OV or V(...)O, or whether in a pragmatically conditioned one we will find rheme-V or V(-...)-rheme#. In such an aporia, then, the analyst's last resort is the statistical approach. However, this can be conditioned by a great number of factors, such as the type of text, the discourse register, perhaps even the idiolect of the author, the parameters scrutinized by the analyst, the methodology employed, and so on.

De Rijk (1998 [1969]) analyzed three samples of texts, the longest one consisting of folktales gathered by a group of ethnologists during the period 1920-1936. In all three of them, the predominant order was SOV. Nonetheless, all six possible permutations among the three elements S, O and V did actually occur in the material, although VOS was extremely rare. On the other hand, in the second sample, consisting of dialogues, the distribution of the different orders was even more extreme than in the other two, and SOV was only slightly more frequent than SVO (cf. de Rijk 1998 [1969]: 16-17).

---

spread in historical Basque has been thoroughly researched by Aldai (2011a). One of his interesting conclusions is the following (2011a: 1129):

"Short verbs favored adjacency, while longer verbs disfavored it. I had also argued in Section 3.3.10 that this correlation between verbal length and adjacency may be predicted on account of Hawkins' principle of MiD [= Minimize Domains], and more explicitly as a result of minimal domain recognition and production. If the verb is shorter than its arguments, domain recognition will be more efficient in the linear order WhVS or WhVO. Thus, adjacency is predicted. Conversely, if the verb is longer, domain recognition will be more efficient with WhSV or WhOV orders. Consequently, nonadjacency is predicted".

De Rijk's view is actually the standard one among Bascologists. However, in the 1990s, Hidalgo (1994, 1995) came up with a series of interesting objections to the standard view. According to Hidalgo, the SOV order of Basque has been considerably favored by the prescriptive work of *anti-erderist* grammarians,[5] i.e. those who have highlighted —and promoted— the features of Basque which are opposed to those of Spanish and French. With regard to word order, this tradition would begin with Cardaberaz (G, 18th century), and reach its peak with Azkue and Severo Altube (B, first half of the 20th century). Since the typical word order in neighboring Romance is SVO, these Bascologists would have overemphasized the tendency of Basque towards SOV, and, their works being bedside books of many Basque (wo)men of letters, they would have inevitably influenced much of the later literature. Previous old literature in Basque, as well as popular speech up to the present day, clearly favor SVO over SOV, in Hidalgo's view.

Although Hidalgo may have overestimated the influence of those grammarians on later authors, the debate sparked by him has posed some very penetrating and hitherto unexplored questions as to what factors favor SOV or SVO. However, as was already pointed out by de Rijk (1998 [1969]), typological arguments support the alignment of Basque with SOV languages. In particular:

1) Historical Basque is an ergative language, and there are extremely few ergative languages with an unmarked SVO order (see Dixon 1994: 50).
2) Historical Basque has copious suffixing, but very little (and somewhat unnatural) prefixing, a feature also related to SOV languages.[6]
3) Historical Basque has the order NPo and GN (see however §10.3.3). These features —particularly the former— are also closely related to SOV.

In short, although statistical data are controversial, historical Basque undoubtedly matches almost all features of SOV languages, and can hardly be defined otherwise.

5 *Erdera* (*erdara* in conservative dialects; cf. *OEH*: '1 erdara') means any language which is not Basque, typically Spanish or French.

6 This argument alone would not be conclusive, since in general, suffixing is less marked than prefixing. There even exist VO languages which are exclusively suffixing, although the ratio of exclusively suffixing languages within VO languages is lower than that of exclusively suffixing languages within OV languages (cf. Hawkins & Gilligan 1988).

10.3. Word order within the Noun Phrase

Since the typological work on word order of the 1970s (particularly Lehmann 1973 and Vennemann 1974), it is usual to group word order parameters together, in such a way that their two possible different values are set out in two columns, corresponding to the general orders Modified-Modifier or Modifier-Modified. This traditional division is represented by Harris & Campbell (1995: 196), as in Table 10.1.

Table 10.1. Some typical Modifier-Modified vs. Modified-Modifier relations

Modifier-Modified	Modified-Modifier
OV	VO
NPo	PrN
AdjN	NAdj
GN	NG
DetN	NDet
RelN	NRel
VAux	AuxV
StAdj	AdjSt
[Harris & Campbell (1995: 196)]	

It is a well known fact, however, that most languages, if not all, do not homogeneously adjust to one of the columns for all their parameters, but show some split. At most, we can speak of tendencies. Basque, for instance, is prone to the Modifier-Modified type in most parameters, but presents some irregularities, as will be shown. According to Comrie (1989 [1981a]: 130), these splits are a problem when ascribing a type to a particular language. However, they seem rather to be a clear indication that such an ascription cannot be made on exclusively synchronic terms without taking into account the evolution of the language. In fact, every language is in constant change, and this change can bring about variations which affect certain specific parameters without affecting others. At least since Sapir's works, we know that languages do not change aimlessly and by chance. Instead, they usually develop in a certain direction, although not all of their elements do so at the same time nor at the same pace. As a result, if we pick a synchronic moment of the language, we will most probably find, side by side, innovating features and archaisms.

As was shown in §10.2.3, almost all Bascologists accept that Basque is an SOV language, and hence that it should be related to the left-hand column of Table 10.1. However, some features diverge from that column. The parameters OV / VO and VAux / AuxV have already been commented on in §10.2. Here, I shall focus on the word order within NPs and adpositional phrases, moving from

those most consistent with the general Modifier-Modified order to those with diachronic/diatopic variation, to arrive at the AdjN vs. NAdj parameter, which in Basque seems to align with the Modified-Modifier order.

10.3.1. NPo vs. PrN

Historical Basque has been exclusively postpositional. The noun governed by the postposition can appear in the bare root form, as in (19a, b), or in the genitive, as in (19a', b'):

(19)	a.	*mahai*	*gain-ean*	a'.	*mahai-aren*	*gain-ean*
		table	top-INES.SG		table-GEN.SG	top-INES.SG
		'On the table'			'On the table (a definite one)'	
	b.	*ohe*	*azpi-an*	b'.	*ohe-aren*	*azpi-an*
		bed	under-INES.SG		bed-GEN.SG	under-INES.SG
		'Under the bed'			'Under the bed (a definite one)'	

In fact, were it not for the locative semantics of all these examples, it would be better to consider (19a, b) as word compounds rather than as NPo-s, and (19a', b') as GNs rather than as NPo-s. Some other meanings usually hosted in postpositions, such as 'in', 'to', 'towards', 'up to', 'from', etc., have become cases, although it is not easy to establish precisely which of them come from postpositions, and how. An undisputed case of a postposition which has entered the declensional system is the motivative suffix, expressed by the postpostion *-ga(i)ti(k)* (cf. §6.5.2.2). But it has done so as a grammatical case (forming the abs. sg. upon the article, as in *gizon-a-gatik* 'because of the man'). In short, from a synchronic point of view, it is not always easy to establish which items should, and which items should not, be defined as postpositions. On the other hand, diachronically, the typical grammaticalization path 2nd member of compound → postposition → case marker must be handled with care, because the processes for each of the cases may have occurred in different ways.[7]

All in all, it seems safe to state that Basque is NPo (Modifier-Modified, harmonic with OV), because every (possible) adpositional construction of the language is NPo, not PrN.[8]

7 See however Trask (*HB*: 246) and Lakarra (2005: 423). Both defend the postpositional origin of e.g. the ergative (cf. §6.4.1.1).

8 Although the parameter PreN vs. NSuff stands outside the word order proper, it is worth recalling that Basque is also almost exclusively suffixing. Actually a few prefixes do

10.3.2. DetN vs. NDet (Det = Article, Demonstrative, Numeral, Interrogative)

Within the category of determiners, Basque articles, demonstratives and the numerals 'one' and, in the B. dialect, 'two', appear to the right of the modified noun, with the restrictions that will be indicated below. The Basque article is a bound morpheme upon which the singular of the noun declension is built, whereas demonstratives and numerals can appear as independent pronouns. Since the three of them show this structural similarity, and since the article comes from a demonstrative, I will include the NArt construction in this group. This is despite the fact that the impossibility of the article *-a / -ak* appearing as an independent word would make it controversial, depending on the analyst, to consider NArt a NDet structure.

Basque has the definite article *-a* for the singular, *-ak* for the plural, and in addition a third one, *-ok*, which is a kind of plural of proximity still productive in western dialects (but also present in old eastern texts). The origin, at least of the singular *-a*, is clear: < **har*, an old 3rd grade demonstrative which in the cases other than the abs. still survives as such (and in B also in the abs.) (Manterola 2006). In old texts and some current varieties, there are traces of old 1st and 2nd grade demonstratives which have been cliticized to the stem (cf. §6.3.1). As for demonstratives (both determiners and pronouns), they can be of three grades: 1st grade *hau / hauek*, 2nd grade *hori / horiek*, 3rd grade *hura* / *haiek* (cf. §7.2.1).

Both articles and demonstrative determiners are located to the right of the modified noun:

(20) a. *mahai-a*
table-ART
'The table'

exist, but they are either loans (*egin* 'to do' → *des-egin* 'to undo'), or loan translations (*egin* → *berr-egin* 'to redo'), and they are somewhat unnatural. Exclusive suffixing increases the degree of closeness to the OV type, although there are also some VO languages which have it (but not the other way round; cf. Hawkins & Gilligan 1988). Let it be said, in passing, that although Standard Basque or *batua* has abundant derivational suffixes, many of these are not commonly used in the spoken language. Besides, the abundance of derivational suffixes decreases as we move towards the 16th century in the text tradition (cf. Bueno 2006). Perhaps this increase in derivational suffixing can be interpreted as another strengthening phenomenon in the general drift towards SOV.

b.	*mahai*	*hau*	b'.	*mahai*	*hori*	b''.	*mahai*	*hura*
	table	this		table	that		table	that there
	'This table'			'That table'			'That table there'	

In Biscayan, we find abundant examples with the demonstrative determiner also to the left. Corresponding to (20b, b', b''), we would have (21a, b, c):

(21) a. *au ma(ha)i au*
b. *ori ma(ha)i ori*
c. *a ma(ha)i-a*

As can be observed, this occurs with all three grades of demonstratives. Until recently, in such constructions, the demonstratives on the left and on the right had to be of the same grade, but it seems that this restriction is becoming increasingly less rigid, so that nowadays it is possible for the 1st or 2nd grade demonstrative to be on the left, and the article on the right, as in (22a, b):[9]

(22) a.	*au*	*ume-a*	b.	*ori*	*ume-a*	
	this	child-ART		that	child-ART	
	'This child'			'That child'		

That these are innovations seems to be supported by the incontrovertible fact that the article comes from the grammaticalization of a 3rd grade demonstrative to the right of the noun. There have also been attempts to grammaticalize as articles the 1st and 2nd grade demonstratives in the same manner (cf. §7.2.4).

With regard to numeral determiners, *bat* 'one' —which in fact should be described as the indefinite article, and hence translated as 'a'— appears to the right of the modifed noun. Leaving *bi* 'two' aside for the moment, from *hiru* 'three' onwards all numeral determiners appear to the left:

(23) a.	*txakur*	*bat*				
	dog	a				
	'A dog'					
b.	*hiru*	*txakur*	b'.	*ehun*	*txakur*	
	three	dog		one hundred	dog	
	'Three dogs'			'One hundred dogs'		

9 I must thank Ariane Ensunza for this observation.

The numeral determiner *bi* 'two' is a special case here. In general, *bi* behaves like any other numeral determiner apart from *bat* 'one': *bi txakur* 'two dogs', etc. However, in the western dialect it appears to the right, as in (24a), and there are traces of the same construction (NP + *bi*) in old eastern attestations, as in (24b):

(24)	a.	*liburu*	*bi*	b.	*esku*	*eta*	*zango*	*bi-ak*
		book	two		hand	and	leg	two-ART.PL
		'Two books'			'The two hands and legs'			
		(Biscayan)			(Oih., *Poems* X, 9th stanza)			

Finally, interrogative determiners are to the left of the noun, in all periods and dialects:

(25)	a.	*Zein*	*gizon*	*etorr-i*	*da?*
		which	man	come-PFV	AUX
		'Which man has come?'			
	b.	*Zenbat*	*gizon*	*etorr-i*	*da?*
		how many	man	come-PFV	AUX
		'How many men have come?'			

In spite of the position of interrogative determiners, Basque is more NDet than DetN, while the direction of the sole attested changes appears to be NDet → DetN (that of the numeral 'two' and perhaps the development *ume-au* → *au ume-au*). According to Table 10.1 above, this would imply that this parameter preserves the old VO related value, and has not moved a long way in the Modified-Modifier → Modifier-Modified direction. However, as Harris & Campbell (1995: 196) themselves recall, it is controversial as to whether this parameter can be correlated to the others in the OV / VO dichotomy, and most scholars leave it out of the set of correlations. I will proceed in the same manner, and in what follows I will dispense with this parameter in the OV / VO discussion. What has been said in §10.3.2, then, must be interpreted in purely descriptive terms.

10.3.3. GN vs. NG

In historical Basque, there are two genitives. One is the possessive genitive, expressed by the suffix *-(r)en*, its old form being *-(r)e*, as in *Ortire semea* 'Orti's

son' (Navarre, 1125), in genitive pronouns like *hire* 'your (sg.)', etc. (cf. §6.4.1.2). The other is the local genitive, expressed by the suffix *-ko*. In historical Basque, the usual position of both genitives is to the left of the modified NP, as in (26a, b):

(26) a. *aita-ren* *etxe-a*
father-GEN.SG house-ART
'The father's house'

b. *herri-ko* *etxe-a*
village-L.GEN.SG house-ART
'The house of the village (= the town hall)'

This fits into the Modifier-Modified order, predominant in OV languages. However, in old texts and some archaic structures, we can find abundant attestations of the local genitive to the right of the modified NP, i.e. of NG structures. Here are some examples:

(27) a. *Milia* *Lastur-ko* (Dirge gathered in the 16th century; cf. *TAV*: 75)
p.n. place.name-L.GEN.SG
'Milia from Lastur'

b. *Peru* *Leartza-ko* (Basque novel by Txillardegi, published in 1960)
p.n. place.name-L.GEN.SG
'Peru from Leartza'

c. *Orazione* *igande-ko* (Prayer contained in Dch. I)
prayer sunday-L.GEN.SG
'Sunday's prayer'

d. *Jaun* *Zeru-ko-a*
lord heaven-L.GEN.SG-ART
'Heaven's Lord'
(Laz., <1152r>)

d'. *Gaztelu* *Gebara-ko-a*
castle p.n.-L.GEN.SG-ART
'The castle of Guevara'
(Laz., <1159r>)

There are also attestations of local genitives to the right of the modified noun in medieval documents, most of them in surnames of toponymic origin. In the documentation of the Navarrese monastery of Iratxe, for instance, we find *Peidro Soria-co* 'Peidro from Soria' (1213), *Garcia Baçterr-eco* 'Garcia from

the Corner' (1227), *Orti Çufico* 'Orti from the Bridge' (13th century). This does not mean that the NG structure is the only possible one in this period. GN also appears in attestations as early as *Sarri-co-huri* 'city of the undergrowth' (*CSM*, 1025), *behen-go landa* 'the land from the bottom' (Tafalla, 1177), *Bazterr-eco ardançea* 'the vineyard of the corner' (Iratxe, 1164).

If we move to the Modern Era, Basque family names containing a GN structure, such as *Aurre-co-echea* 'the house from the front part / in front', *Urruti-co-echea* 'the house in the distance', *Goi-co-echea* 'the house from the upper part', etc., are attested from at least the 16th century. But so are some others, apparently more archaic, which appear to be NG in their structure, despite not having *-co*: *Echa-guibel* 'the house of the back part', *Echa-guren* 'the house of the highest part', *Echa-mendia* 'the house of the mountain', etc.

As far as the local genitive is concerned, then, both the GN and the NG structures seem to have existed side by side, until GN swept NG almost completely aside. Perhaps the two co-existed even in Common Basque. But for reasons that we will consider in §10.4, the best candidate for the original one is NG, the one that is harmonic with the Modified-Modifier order and hence with VO languages.[10]

10.3.4. RelN vs. NRel

Basque can build relative clauses in several ways. The most typical one consists in adding *-(a/e)n* to the finite verb, that is, to the only verbal form if this is synthetic (28a), or to the auxiliary verb in periphrastic verbs (28b):

(28) a. *dauka-da-n* *liburu-a*
have-1SG-RLV book-ART
'The book that I have'

10 NG structures with the possessive genitive are rarer, even in old texts and documents, and mainly appear in patronymic family names such as *Guarcea Ansorena*, *Sanso Urraquarena* (both in Iratxe, Navarre, 13th century, although this patronymic structure with the suffix *-(r)en-a* has existed to the present day). Some others are found in verse, such as Dch. *Lehen inprimazaili-a heuskara-ren hura da* 'The first printer of Basque is he', E.Cib. *Errege Iudu-en-a* 'the King of the Jews', *furia Satan-en-a* 'Satan's fury', and in these cases metrical reasons could be involved. Ulibarri (2010: 102-103) has also noticed the existence of at least two NG structures with possessive genitive in the VJ, as in *tentaziñoe geure arerio-en-etarik* 'from the temptations of our enemies'.

b. *ikus-i* *du-da-n* *gizon-a*
see-PFV AUX-1SG-RLV man-ART
'The man that I have seen'

The relative clause, then, precedes the NP modified by it, so that we can classify this structure as being of the RelN type, typical of OV languages. This is the word order of relative clauses in the present day. However, we can find at least three types of traces of an NRel order for the same structure, which is apparently older.

The first type of trace, already pointed out by de Rijk (1998 [1980]), is that of NRel structures in old texts of any dialect, as in (29a, b), and in far eastern peninsular dialects until the 20th century, as in (29c):

(29) a. *Zure* *pena* *di-o-zu-n-ok* (Dch. IX, 30)
your grief say-3SG-2SG-RLV-ART.PL.PROX
'Those griefs of yours that you talk about'

b. *Sakramentu* *artu-ko* *do-gu-z-(a)n-ak* (Cap., *DC*)
sacrament take-PROSP AUX-1PL-3PL-RLV-ART.PL
'The sacraments that we shall take'

c. *Agin* *min* *emoi-ten* *dai-a-n-a* (R, ca. 1900)
tooth pain give-IPFV AUX-2SG-RLV-ART
'The back tooth which hurts you'

The second type of trace was pointed out by Lakarra (2005: 422), and is to be found in some words of the Basque common lexicon, such as *urdin* 'blue', *berdin* 'same', *bizardun* 'bearded man'. Michelena (1987 [1970]: 290) already noted that these seem to be fossilized relative clauses with the verb *-din* 'which has become' and *-dun* 'which has'. They would come, respectively, from something close to 'which has become water', 'which has become the same', 'who has a beard'. Since all these are typically adjectives and hence postponed, the NP that they form with the preceding noun constitutes an NRel structure, as in *ile ur-din* 'hair which has become water-like'.[11]

11 However, Lakarra (p.c.) has pointed out to me that this argument might weaken if we consider —as in turn de Rijk (p.c.) pointed out to him— that, had *-din* and *-dun* been finite verbs of Rel clauses, then we should start from **da-din-en* and **da-dun-en* in CB, and we would expect dialectal variation in the phonetic evolutions of these proto-forms. If, by contrast, we start from *-din* and *-dun*, then the historical *gor-din*, *bizar-dun*, etc.,

The third type of trace of NRel structures is a series of toponymic *Satznamen*, some of them gathered in Kontrasta (east of Alava), such as *Aiz-luceadana* 'crag which is long' (Kontrasta, 1787) and *Ur-diraquiana* 'water which boils' (Amescoa, western Navarre). They had already been studied by Michelena (1988 [1985]), but have been included in this group of NRel structures by Lakarra (2005: 422).

In view of all these data, it seems highly probable that, over the history of Basque, a drift from NRel (Modified-Modifier, harmonic with VO) to RelN (Modifier-Modified, harmonic with OV) has slowly taken place. Although the latter has prevailed, there remain several traces of the former.

10.3.5. AdjN vs. NAdj

This parameter differs from all the others, for the value displayed by Basque throughout all of its periods and dialects is NAdj, which is Modified-Modifier and harmonic not with OV, but with VO. Since Basque is a canonical SOV language in other respects, we should expect to find AdjN, but this order is actually ungrammatical.[12] This is shown in (30a, b):

(30)	a.	*etxe*	*gorri-a*	b.	*mutil*	*handi-a*
		house	red-ART		boy	big-ART
		'The red house'			'The big boy'	

Greenberg (1963) had already noted this irregularity of Basque, but disregarded it as an unimportant phenomenon, since such inconsistencies exist in many other languages (as in English, AdjN but VO in most other respects).

However, as Michelena (1990 [1963]: 11) recalled, from a diachronic point of view, a synchronic irregularity of a state of the language is a reflection of a previous regularity, erased or blurred over time. This observation has been taken to the extreme by Lakarra (2005: 420; cf. also §5.6). According to him, although in synchronic terms the NAdj order of Basque may only be an anecdote, a diachronist cannot obviate that feature. Instead, it is precisely in this feature

---

can be explained, but they would no longer originally be Rel clauses in NRel constructions, but compounds of the structure 'noun – verbal root'.

12 Exceptionally, adjectives with the suffix *-(t)ar*, which in most cases indicates geographical origin, as in *bizkai-tar* 'Biscayan', or propensity, as in *eskuin-dar* 'right-winger', can be located either to the left or to the right of the noun. But this does not seem to be significant to our diachronic problem, or at least nobody has introduced it as a new element in this issue.

that (s)he should locate the oldest state of affairs, and search for further traces of the language which are harmonic with that exceptional feature. As he notes, "nothing has been achieved in reconstruction starting from regularities" [translation by the editor, MMA].

Furthermore, Lakarra (2005: 419) relates this "irregularity" to another "curiosity" of Basque: there are reasons to doubt that the Adj has always been a large and open lexical category in the history of Basque, clearly distinct from Ns and Vs.[13] An indication in favor of this view is that color names are clearly derived (*beltz* 'black' < *bel-tz*), participles (*zuri* 'white' < *zur* 'wood' + *-i* 'PTCP', *gorri* 'red' < *gor* 'deaf, naked' + *-i* 'PTCP'), relative clauses (*urdin* 'blue' < *ur* 'water' + *-(da)din* 'become, turned into'), or loan words (*berde* 'green' < Sp. *verde*, *marroi* 'brown' < Sp. *marrón*). Lakarra does not believe that these two features —the position of the Adj to the right of the N, and the lack of an open adjectival class in the old language— are anecdotal, nor independent from one another. On the contrary, according to him, since the parameter of Adj did not at first exist as such, it is only natural that, when it arose, it occupied the position to the right of the noun, as that position is the one corresponding to at least three of the four sources from which adjectives come.

Finally, according to Lakarra (2006b: 270) the NAdj order also provides some support for Gómez's view (cf. §10.4.2) that the old word order is VSO, as in this type the NAdj order is usually obligatory, whereas in SVO and SOV languages the opposite order AdjN also occurs.

10.4. Hypotheses about the word order in Proto-Basque

Now that the basic word order phenomena of Basque have been described, I shall proceed to summarize the main theories about the word order in Proto-Basque. It is important to note that this must be previous to Common Basque. Now, since all Basque dialects are ergative, of unmarked SOV order and NPo, it is reasonable to believe that CB must also have been roughly SOV and NPo. The parameters GN vs. NG and RelN vs. NRel, by contrast, show some degree of variation and diachronic dynamicity throughout their attested history. As has been argued, the direction in which these parameters have been moving is Modified-Modifier → Modifier-Modified, and hence we can only expect that CB was more Modified-Modifier than historical Basque with regard to these parameters, although it is impossible to say to what degree.

13 In fact, this lack had already been noted by Humboldt.

However, most diachronic analyses of Basque word order have focused on stages which are to be located prior to CB, as they involve changes in the structure of common words, especially of common finite verbs. Consequently, these analyses resort to internal reconstruction. According to all of these analyses, there are solid grounds to believe that, at some stage previous to CB, Proto-Basque was VO. Nonetheless, they disagree as to what VO order —mainly SVO or VSO— is to be preferred. Let us look at some of these analyses.

10.4.1. Trask (1977, HB*)*

Trask (1977) was the first attempt at a reconstruction of the word order of PB carried out by analyzing the order of elements in the finite verb. Based on the idea that today's morphology is yesterday's syntax, he noted that in intransitive finite verbs of present conjugations beginning with 1st and 2nd person morphemes, the order of elements corresponds to SV (Subject-Verb). Thus, he (1977: 206) reconstructed *na-go* 'I am (in/at)' as **na-da-go* or **ni-da-go*, where *-da-* was to be analyzed as an aspect marker. If we add to this the fact that non-finite verbs have at least one element which is clearly a fossilized prefix of some stage (*e-*, and prefixes are quite at odds with SOV), then it seems sound to reconstruct SVO as the clause-level order of Basque for some stage.[14] He summarized this hypothesis as follows:

> "[...] I propose that the SVO morphological order of Basque intransitive verbs should be taken as evidence for an older SVO syntactic order. That is, I propose that the originally independent pronouns came to be bound to the verb at a time when the structure of the Basque sentence was SVO, and that when this structure changed to SOV, the pronominal elements, being bound, could not move, but remained morphologically "frozen" in their earlier syntactic position" (Trask 1977: 208).

What about (present tense) transitive finite verbs, in which in historical Basque the preverbal element is not the S but the O (e.g. ***na***-*kar* '(s)he brings **me**')? According to Trask, such forms could also be brought back to an SV structure by simply assuming that these transitive verbs were originally passive, and were later reinterpreted as active. Hence, *na-kar* would come from **na-da-kar* or **ni-da-kar*, the pronominal element being a patient subject in origin. As possible parallels which have undergone similar developments, he mentions some Poly-

14 A third element supporting a previous VO phase according to Trask would be the VO order of negative sentences in historical Basque (cf. (6a) in §10.2.1.2). He recalls that negative sentences are typical archaizing contexts.

nesian and Indo-Iranian languages. He develops some of these arguments in Trask (1979) and Trask (*HB*: 245-247).

10.4.2. Gómez (1994) and Gómez & Sainz (1995)

The view defended in these papers also starts from the assumption that today's morphology reflects yesterday's syntax, but a different conclusion is reached. The theory defended here is that when clitic pronouns became morphemes bound to the verb, Basque was a V-initial language, either VSO or VOS. However, as in many other V-initial languages, the V did not stand in the absolute initial position. Instead, there was always an element which prevented the V from being the first element in the sentence, so that the order was #X-V-.

When the abs. marker was overt, as is the case with 1st and 2nd person morphemes, it was the clitic representing it that occupied the position of X: ABS-V(-DAT)-ERG (as in *na-kar-zu* 'you are bringing me'). When, on the contrary, the abs. was a 3rd person, and hence covertly marked as *Ø-*, some other marker had to prevent the verb from standing sentence-initially. In the (historically) present tenses, that marker was the particle *da-*, as in **da-du-da* (> *dut* 'I have'), i.e. da-V-ERG, or in *da-tor* '(s)he comes', i.e. da-V. The same would hold for *le-* with irrealis tenses, and for *be-* with imperative ones.

In the (historically) past tenses, if the abs. is a 3rd person, 1st and 2nd person erg. clitics are moved to the X position, and the order is ERG-V (*ne-karr-en* ERG-V- 'I brought'). As in the present, 1st and 2nd abs. clitics are in the X position. If both the abs. and the erg. are 3rd persons, no overt pronominal clitic is at hand. The morpheme which appears, *ze-* (*e-* in the west, in most cases), is related by Gómez (1994: 101) and Gómez & Sainz (1995: 266) to the sentence connective *ezen* 'that', which in this case would be profited to protect the verb from standing at the beginning of the sentence.[15]

The theory as a whole is attractive for the past tense, but in the present tense one problem is how to explain the *-a-* of 1st and 2nd person bound abs. morphemes, as we would expect e.g. ***ni-tor-* instead of *na-tor* to fit into the alleged #X-V- structure. The solution proposed by Gómez (1994: 107) is that the prefixing process took place in two phases. First, *da-* was attached to protect the V from the initial position, > *da-tor*, and in a second phase the abs. clitic was

15 Gómez & Sainz (1995: 268) compare this phenomenon to the insertion of the dummy *no-* in Old Irish to host pronominal infixes as in *nom-chara* 'he loves me'. This *no-* is analyzed by Watkins as related to the Hittite connective *nu*. In Old Irish too, this connective was reanalyzed as a third person marker.

prefixed to *da-*. The phonetic merger of the two gave rise to the person agreement markers.

10.4.3. Lakarra (2005, 2006b)

Based on the diachronic holistic typology developed by Donegan & Stampe (1983, 2004), Lakarra (2005: 449-454, 2006b: 278-282) has proposed that Basque has undergone a holistic or thorough typological drift which has displaced it from one end of the typological spectrum (VO) to the opposite one (OV). This would also coincide with a change from an isolating structure to an agglutinating one.

The holistic typology proposed by Donegan & Stampe is an attempt to integrate more parameters into the framework already set out by Lehmann (1973) and Vennemann (1974). In the latter, the parameters which made up the typological frame were exclusively those related to word order (the ones discussed for Basque in §10.2 and §10.3). By contrast, in their analysis, Donegan & Stampe include parameters such as type of morphology, syllable canon, accent/prosody, timing, type of consonant-system, type of vowel-system, etc. The different values for each of these parameters are also binarily classified, and assigned either to the OV type or the VO type. Table 5.2 of Lakarra's chapter in this volume sets out the most representative parameters (and values) considered by Donegan & Stampe.

That this holistic approach is dynamic and can provide diachronic explanations for an array of phenomena in different languages or language families is exemplified by comparing the Mon-Khmer languages to the Munda ones. While the former ones have a rising phrase accent, rigid SVO and NAdj orders, are prepositional, analytic, fusional, (mostly) prefixing, isoaccentual, (C)V- or (C)(C)´V(C)(C) for syllable canon, etc., the latter, by contrast, have a falling phrase accent, variable SOV and AdjN orders, are postpositional, have case and agreement systems, are agglutinative, suffixing, isosyllabic or isomoric, (C)V(C) for syllable canon, etc. Disregarding substratic accounts to explain such a drastic contrast, since both groups are traditionally genetically related anyway, Donegan & Stampe suggest that Mon-Khmer is archaic in most respects, and has inherited unchanged the features of Proto-Austroasiatic, whereas Munda has undergone a holistic drift towards the OV type which has turned all the values for every parameter, originally harmonic with the VO type, into the opposite ones.

Lakarra suggests a similar holistic drift in the development from Proto-Basque to contemporary Basque. According to him, even though nearly all the values of today's Basque are OV-harmonic, for practically every value there are

traces of a stage in which each of those values was VO-harmonic. As has been shown in Section §10.3, for some word order parameters such as GN vs. NG and RelN vs. NRel, the drift from VO to OV is still underway in the historical period. The exception is NAdj, maintained as such to the present day, which entails an irregularity from the synchronic point of view, but a remnant of the original situation from a diachronic one.

Lakarra (2005: 420) also concludes that, since there are grounds to believe that before the holistic drift Proto-Basque was prepositional and highly prefixing (cf. §10.3.5),[16] since Basque has always been exclusively NAdj, and since exclusive PrN, a high tendency towards PreN and exclusive NAdj are features closely associated with the VSO order[17] —among the three most usual orders, the one located furthest to the left along the VO-OV axis— VSO is precisely the order that ought to be favored in reconstructing work, as proposed by Gómez (1994) and Gómez & Sainz (1995).

10.4.4. Martínez-Areta (2011, 2012)

Martínez-Areta (2011, 2012) is a refinement of Trask (1977), also compatible with —or even complementary to— Lakarra's (2005) drift from VO to OV (although it is actually incompatible with any V-initial order). According to Martínez-Areta, the whole of the present verbal morphology comes not from morphological passives but rather from passive interpretations of constructions

16 In Lakarra's model of root structure, lexemes are CVC. Thus, in a common pattern of patrimonial words like *sa-bel* 'stomach', *sa-kon* 'deep', *sa-min* 'bitter', *gi-bel* 'liver', *gi-har* 'lean meat', etc., the first elements *sa-* and *gi-* cannot be lexemes, and hence they must be prepositions or prefixes. However, the clearest case of abundant prefixing seems to be the morphology of non-finite verbs, brilliantly elucidated by Lakarra himself (cf. §5.7, §8.3.1.2.4). From a verbal root such as *gotz*, for instance, the following derivations can be made to the left: *e-gotz* 'to push, to impel', *e-ra-gotz* 'to prevent', **e-da-ra-gotz* (> *urgatz* 'to assist'). By contrast, morpheme addition to the right does not seem to be so copious. To each of the previous verbs, the participle *-i-*, the imperfective *-te-*, and perhaps some other suffixes, depending on the analyst, can be added, but they do not appear piled on top of one other, or are not of a derivational nature.

17 Obviously, this assertion depends on the criterion adopted with regard to each typological issue, and typologists have often become embroiled in endless discussions, from which it is not easy for the non-expert to reach a reliable conclusion. In the case of AdjN vs. NAdj, it has even been debated as to whether the parameter itself can be correlated to the parameter OV vs. VO and the other word order parameters. Against the correlation, cf. Dryer (1988). In favor of the correlation, cf. Hengeveld & Rijkhoff & Siewierska (2004).

which existed at a stage in which the language had no grammaticalized diathetical opposition, just as in Chinese. Thus, *nakus* '(s)he sees me', probably from **ni da-kus*, would originally mean 'I am seeing' or 'I am being seen', depending on the context, or on whether it was followed by an O (as in **ni da-kus gizon* 'I am seeing the man') or by an agentive complement (as in **ni da-kus gizon(-e)-ga* 'I am being seen by the man'). Since, in all finite verbs of the present conjugations which have 1st or 2nd person bound morphemes at the beginning, these 1st and 2nd person morphemes are patient markers, this means that the whole (historically) present morphology derives from passive interpretations.

In those cases where the patient was a 3rd person, the prevalence of the passive (clause-level) construction would imply that e.g. 'The shepherd is seeing a/some sheep' would not be expressed as **ardizani dakus ardi* (i.e. shepherd - is seeing - sheep), but as **ardi dakus ardizani-ga*, lit. 'A/some sheep is/are being seen by the shepherd' (the example is obviously fictitious). Such a drift towards the passivization of any semantically transitive sentence, however, can bring about a certain degree of discourse interference, since the agentive complement, which is typically animate, definite, and known information, is always at the end of the sentence, whereas the patient, more often inanimate, indefinite, and new information, is always at the beginning. A way of solving this problem can be to topicalize the agentive complement: → **ardizani-ga ardi dakus* 'By the shepherd a/some sheep is/are being seen'. Once this procedure had become general, the topic grammaticalized as (agent) subject and the whole sentence was actively reinterpreted, → 'The shepherd is seeing a/some sheep'. Thereby, the language became ergative, and the change SVO → SOV was accomplished.[18]

18 A kind of preliminary version of this account is already present in Trask (*HB*: 247). However, in line with the typology proposed by Li & Thompson (1976), Martínez-Areta assigns a *topic-comment* discourse structure to PB. With 1st and 2nd person patients, the topicalization, active reinterpretation, and evolution into a *subject-predicate* structure would imply that, say, the sentence **ni dakus ardizani-ga* 'I am being seen by the shepherd' turns into **ardizani-ga ni dakus* 'By the shepherd I am seen', and finally → 'The shepherd is seeing me'. Since here *ni* is no longer the independent subject of a comment but the O of the new predicate, the way is paved for its cliticization and transformation into a bound morpheme, > **ardizani-ga n(i)(-)(d)akus*, thus giving rise to a verbal system with grammaticalized person agreement. In a formula, rather than SVO, what Martínez-Areta proposes for PB is a TSV(O) order (where T = Topic, and where there is no post-verbal O if this is the T).

10.5. Conclusion

In the previous sections, it has been shown that contemporary Basque is best defined as an SOV language, and has features related to the OV order in nearly all parameters, both at clause level and within NPs. It must be recalled, however, that language is dynamic and continuously changes. This brings about splits between OV-related values and VO-related ones, which can eventually lead to a typological drift, although obviously not all aspects change simultaneously and at the same pace. Hence, how that drift has taken or will take place in a given language is highly unpredictable. A synchronic state of the language, then, is not a finished product, but a phase of an ongoing process, like a photogram of a film. As a result, it is difficult to find a language with no split at all, that is, completely OV or completely VO.

In the case of Basque, in practically every parameter there are traces which point to a development in the direction Modified-Modifier → Modifier-Modified, taken to the extreme in Lakarra's holistic theory. The AdjN vs. NAdj parameter seems to be a fossil which has retained the Modified-Modifier value until the present day. The question as to at what point in this process Common Basque was, requires systematic work that has not yet been undertaken. In any case, if we consider that all dialects are SOV, we are obliged to think that the common language was SOV too. This seems more reasonable than assigning it another order and assuming that all dialects have developed their historical structure in an independent way but in the same direction. This obviously does not mean that all the word order parameters of Common Basque were already the same as in historical Basque.

Finally, internal reconstruction gives us grounds to think that there existed, at some point in time before Common Basque —which we could call Proto-Basque—, a VO consituent order, whether it be SVO or VSO.

The two main tasks which future word order studies ought to tackle, then, are: 1) to what extent Common Basque differed from historical Basque with regard to some parameters such as RelN vs. NRel, GN vs. NG, or even OV vs. VO, considering factors such as the length of the O and of the V; and 2) to what previous order in Proto-Basque, internal arguments, with the help of cross-linguistic typological research, can lead us. The path is open, then, for further and better studies.

References

Abaitua, J. & Trask, R. L. (1987): "Accusativity in Basque: reply to Bossong". In: *Linguistics*. Nr. 25. P. 395-401.

------- & Unzueta, M. (2011): "Ponderación bibliográfica en historiografía lingüística. El caso de la «vasconización tardía»". In: *Oihenart*. Nr. 26. P. 5-26.

Agud, M. & Michelena, L. (eds.) (1958): "Introducción [al diccionario de N. Landucci]" (Supplements of *ASJU* 5). Donostia-San Sebastián [in *SHLV*. P. 762-782].

------- & Tovar, A. (1988-93): *Materiales para un diccionario etimológico de la lengua vasca (A-orloi)* (Supplements of *ASJU* 13, 19, 24, 26, 30, 33, 37). Donostia-San Sebastián.

Aikhenvald, A. Y. (2006): "Serial verb constructions in typological perspective". In: Aikhenvald, A. Y. & Dixon, R. M. W. (eds.). P. 1-68.

------- & Dixon, R. M. W. (eds.) (2006): *Serial Verb Constructions: A Cross-Linguistic Typology*. Oxford University Press.

Alberdi, J. (1995): "The Development of the Basque System of Terms of Address and the Allocutive Conjugation". In: Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.). P. 275-293.

------- (2011): "Euskararen tratamenduak eta aditz alokutiboa". In: Salaburu, P. & Goenaga, P. & Sarasola, I. (eds.).

Aldai, G. (1994): "Euskal aditz zaharraz: aspektu oposaketatik modu oposaketara". In: *Enseiukarrean*. Nr. 10. P. 67-84.

------- (1999): "Hegoaldeko Goinafarrera zaharra: aditz laguntzaileekiko ezaugarri zenbait". In: *FLV*. Nr. 81. P. 263-274.

------- (2000): "Split ergativity in Basque: the pre-Basque antipassive-imperfective hypothesis". In: *Folia Linguistica Historica*. Nr. 21. P. 31-98.

------- (2003): *Grammaticalization of present and past in Basque*. PhD presented at the University of Southern California.

------- (2006): "Lehenaldi perfektiboko formak XVI. mendeko euskaran". In: Lakarra, J. A. & Hualde, J. I. (eds.). P. 23-34.

------- (2008): "Sobre el significado arcaico de las formas con prefijo *l-*: su uso en Lazarraga". In: *FLV*. Nr. 108. P. 301-327.

------- (2011a): "*Wh*-questions and SOV languages in Hawkins' (2004) theory: Evidence from Basque". In: *Linguistics*. Nr. 49. P. 1079-1135.

------- (2011b): "Sobre el origen de Martín Portal, autor de la poesía premiada en Pamplona en 1610". In: *FLV*. Nr. 114. P. 93-115.

Altuna, P. (1978): "Erakusleak zenbait idazle zaharrengan". In: *Euskera*. Nr. 24. P. 699-712.

------- (ed.) (1996): *Federiko Krutwig-i omenaldia* (*Iker* 10). Bilbao: Euskaltzaindia.

Allières, J. (1981): "La dialectologie basque". In: *Euskal linguistika eta literatura: Bide berriak*. Bilbao: University of Deusto. P. 103-113.

Amonarriz, K. (2008): "Gazte hizkerak hizkera gazteak". In: *BAT Soziolinguistika Aldizkaria*. Nr. 68(3). P. 167-183.

Amorrortu, E. & Ortega, A. (2009): "Zergatik eskolaratzen zituzten gurasoek seme-alabak euskarazko ereduetan?". In: *Euskara*. Nr. 54. P. 579-602.

------- & ------- & Idiazabal, I. & Barreña, A. (2009): *Actitudes y prejuicios de los castellanohablantes hacia el euskera*. Vitoria-Gasteiz: Basque Government.

Anderson, G. D. S. (2006): *Auxiliary Verb Constructions*. Oxford University Press.

Anderson, J. M. & Jones, C. (eds.) (1974): *Historical Linguistics I: Syntax, Morphology, and Internal and Comparative Reconstruction*. Amsterdam: North-Holland.

Apalauza, A. (2008): *Imozko euskara*. Pamplona-Iruñea: Government of Navarre.

Apraiz, O. de (1920): "Un caso de fonética histórica estudiado en la toponimia alavesa: la alternancia *-l-* / *-r-*". In: *RIEV*. Nr. 11. P. 81-94.

------- (1930): "El idioma en Álava. Toponimia alavesa. Notas, problemas y deducciones". In: *Euskal esnalea*. Nr. 20. P. 13-18.

------- & Knörr, K. (1978): "De la toponimia euskariana en Álava". In: *Boletín de la Institución Sancho el Sabio*. Nr. 22. P. 286-304.

Arana, A. (2006): *Nafarroako euskalkiak. Zaraitzuera. Ikasliburua*. Donostia-San Sebastián: Hiria.

Arbelaiz, J. J. (1978): *Las etimologías vascas en la obra de Luis Michelena*. Tolosa: Kardaberatz.

Arejita, A. (ed.) (1988): "Amilletaren Dotrinea. Textoa eta gramatika-oharrak". In: *Litterae Vasconicae (Euskararen Iker Atalak)*. Nr. 4. P. 97-128.

------- (introd.) (2008): *Dotrina christiñauena (by Ochoa of Capanaga, translated from Geronimo Ripalda's Exposición breve de la doctrina christiana)*. Bilbao: Labayru Ikastegia.

------- *et alii* (eds.) (2002): *Bilbao. El espacio lingüístico. Simposio 700 Aniversario*. Bilbao: University of Deusto.

Ariztimuño, B. (2011): *Antzinateko euskal idazkunak eta aitzineuskararen berreraiketa: corpusa, analisia eta ondorioak*. Master Thesis presented at the University of the Basque Country.

------- (2012a): "Etimologiak ate-joka: *din-don...* (*jo* eta *josi*; *eman* eta *iramo(n)*; *jin* eta *jainko*)". In: *FLV*. Nr. 114. P. 119-128.

------- (2012b): "The origin of the Basque partitive". In: Luraghi, S. & Huumo, T. (eds.) (in press).

------- & Egurtzegi, A. (2011): "Enarak, elaiak eta testuinguruaren garrantzia". In: *ASJU*. Nr. 45. P. 403-410.

Arkotxa, A. & Oiharçabal, B. (2009): "Siglo XVI: las primicias de las letras vascas" [at http://www.basqueliterature.com/basque/historia/klasikoa/XVI.%20mendea, accessed on 2011-10-23].

Arriolabengoa, J. (2008): *Euskara Ibargüen-Cachopín kronikan*. Bilbao: Euskaltzaindia.

Artiagoitia, X. (1990): "Sobre la estructura de la sílaba en (proto)vasco y algunos fenómenos conexos". In: *ASJU*. Nr. 24. P. 327-350.

------- (1993): "Syllable Structure in Basque and Proto-Basque". In: Hualde, J. I. & Ortiz de Urbina, J. (eds.). P. 263-287.

------- & Goenaga, P. & Lakarra, J. A. (eds.) (2002): *Festschrift for Rudolf P. G. de Rijk* (Supplements of *ASJU* 44). Bilbao.

------- & Lakarra, J. A. (eds.) (2008): *Gramatika jaietan Patxi Goenagaren omenez* (Supplements of *ASJU* 51). Bilbao: University of the Basque Country.

Arzamendi, J. (1985): *Términos vascos en documentos medievales de los ss. XI-XVI*. Bilbao: University of the Basque Country.

Astarloa, P. P. (1988 [1803]): *Apología de la lengua bascongada, o ensayo crítico filosófico de su perfección y antigüedad*. Bilbao: Editorial Amigos del Libro Vasco.

Aulestia, G. (ed.) (1996): *Los escritores, hitos de la literatura clásica euskérica*. Vitoria-Gasteiz: Fundación "Sancho el Sabio".

Aurrekoetxea, G. (1994): "Hitanoa ahozko hizkuntzan (Euskal Herriko Hizkuntz Atlasaren arabera)". In: *Euskera*. Nr. 39. P. 995-1007.

------- & Videgain, X. (2005): *"Bourciez" bildumako euskal atlasa (BBEA): 2. Gramatika* (= *ASJU* 29(1)).

------- & Ormaetxea, J. L. (eds.) (2010): *Tools for Linguistic Variation* (Supplements of *ASJU* 53). Bilbao: University of the Basque Country.

Azkarate, A. (1999): *Necrópolis tardoantigua de Aldaieta. Vol. I. Memoria de la excavación e inventario de los hallazgos de Gamboa (Álava)* [at http://www.ehu.es/arqueologiadelaarquitectura/PDFs/Aldaieta.pdf, accessed on 2012-09-10].

------- (2004): "El País Vasco en los siglos inmediatos a la desaparición del Imperio Romano". In: Barruso-Barés, P. & Lema-Pueyo, J. A. (coord.). P. 23-50.

Azkarate, M. & Altuna, P. (2001): *Euskal morfologiaren historia*. Donostia-San Sebastián: Elkarlanean.

Azkue, R. M. (1969 [1905-06]): *Diccionario vasco-español-francés* (2 vols.). Bilbao: La Gran Enciclopedia Vasca.

------- (1969 [1923-25]): *Morfología Vasca*. Bilbao: La Gran Enciclopedia Vasca.

------- (1931): "Particularidades del dialecto roncalés". In: *Euskera*. Nr. 12. P. 207-406.

Aztiker (2002): *Euskal Herria. Datuen talaiatik*. Astigarraga: Udalbiltza.

Aztiker and Soziolinguistika Klusterra (2011): [at http://www.soziolinguistika.org, accessed on 2012-01-30].

Bakrò-Nagy, M. Sz. (1992): *Proto-Phonotactics. Phonotactic investigation of the PU and PFU consonant system*. Wiesbaden: Harrassowitz Verlag.

Baldi, P. & Page, B. R. (2006): "Review of *Europa Vasconica – Europa Semitica*". In: *Lingua*. Nr. 116. P. 2183-2220.

Barrena, E. (1989): *La formación histórica de Guipúzcoa. Transformaciones en la organización social de un territorio cantábrico durante la época altomedieval*. Donostia-San Sebastián: University of Deusto.

Barruso-Barés, P. & Lema-Pueyo, J. A. (coord.) (2004): *Historia del País Vasco. Edad Media (siglos V-XV)*. Donostia-San Sebastián: Hiria.

Basque Government (2003): *La continuidad del euskara III*. Vitoria-Gasteiz: Servicio de Publicaciones del Gobierno Vasco.

------- (2008): *IV Encuesta Sociolingüística. 2006. Euskal Herria. Comunidad autónoma Vasca. Navarra. País Vasco Norte. Treviño*. Vitoria-Gasteiz: Servicio de Publicaciones del Gobierno Vasco.

------- (2009) *IV mapa sociolingüístico*. Vitoria-Gasteiz: Servicio de Publicaciones del Gobierno Vasco.

------- (2011): *Language Indicator System of the Basque Country database* [at http://www1.euskadi.net/euskara_adierazleak/indice.apl?hizk=i, accessed on 2012-01-30].

------- (2012): *V. Inkesta soziolinguistikoa. Euskal Autonomia Erkidegoa, Nafarroa eta Iparraldea* [available at http://www.kultura.ejgv.euskadi.net/r46-704/eu 16-07-2012, accessed on 2012-11-14].

Baxter, A. N. (1988): *A grammar of Kristang (Malacca Creole Portuguese)*. Canberra: Pacific Linguistics.

Belasko, M. (2004): "Toponimia y panorama lingüístico de la ribera de Navarra y comarcas circundantes". In: Jimeno, R. & López-Mugartza, J. L. (eds.). P. 55-78.

Bender, M. L. & Takacs, G. & Appleyard, D. L. (eds.) (2003): *Selected Comparative-Historical Afrasian Linguistic Sudies in Memory of Igor M. Diakonoff*. Lincom Europa.

Benveniste, É. (1935): *Origines de la formation des noms en indoeuropéen*. Paris: Maisonneuve.

------- (1969): *Le vocabulaire des institutions indo-européennes* (2 vols.). Paris: Les Éditions de Minuit.

Berria [available at http://www.berria.info, accessed on 2011-03-13].

Bilbao, G. *et alii* (eds.) (2010): *Lazarraga eskuizkribua: edizioa eta azterketa (1.0)*. Vitoria-Gasteiz: University of the Basque Country [available at http://www.lazarraga.com, accessed on 2012-09-02].

Blevins, J. (2004): *Evolutionary phonology: the emergence of sound patterns*. Cambridge University Press.

------- & Garrett, A. (1998): "The origins of consonant-vowel metathesis". In: *Language*. Nr. 74. P. 508-556.

------- & ------- (2004): "The evolution of metathesis". In: Hayes, B. & Kirchner, R. & Steriade, D. (eds.). P. 117-156.

Blust, R. A. (1988): *Austronesian Root Theory. An essay on the limits of morphology*. Amsterdam: John Benjamins.

Böhme, H. W. (2002): "Der Friedhof von Aldaieta in Kantabrien – Zeugnis für ein fränkisches Schlachtfeld des 6. Jahrhunderts". In: *Acta Praehistorica et Archaeologica*. Nr. 34. P. 135-150.

Bonaparte, L.-L. (1862): *Langue basque et langues finnoises*. London.

------- (1866): *Formulaire de Prône en Langue Basque, conservé naguère dans l'Eglise d'Arbonne*. London.

------- (1991 [1869]): *Le verbe basque en tableaux* [in Bonaparte (1991). Vol. 1. P. 175-442].

------- (1991) *Opera omnia vasconice* (4 vols.). Bilbao: Euskaltzaindia.

Brun, A. (1923): *L'introduction de la langue française en Bearn et en Roussillon*. Paris: Honoré Champion.

Bueno, A. (2006): "Atzizkiak aztergai autore ezberdinen begiradapean". In: Lakarra, J. A. & Hualde, J. I. (eds.). P. 221-244.

Bybee, J. L. & Perkins, R. & Pagliuca, W. (1990): "On the asymmetries in the affixation of grammatical material". In: Croft, W. & Denning, K. & Kemmer, S. (eds.). P. 1-42.

------- (1994): *The evolution of grammar: Tense, aspect and modality in the languages of the world*. University of Chicago Press.

Byrne, F. (1990): “Tense marking in serial structures”. In: Joseph, B. D. & Zwicky, A. M. (eds.). P. 149-177.

Camino, I. (1997): *Aezkoako euskararen azterketa dialektologikoa*. Pamplona-Iruñea: Government of Navarre.

------- (2004): *Hego-nafarrera*. Pamplona-Iruñea: Government of Navarre.

------- (2008): “Nafarroa Behereko euskara zaharra”. In: *ASJU*. Nr. 42. P. 101-169.

------- (2011): “Ekialdeko euskararen iraganaz” (to be published in Epelde, I. & Zuazo, K., eds.).

------- (ed.) (1998): *Nafarroako hizkerak. Nafarroako Euskal Dialektologiako Jardunaldietako Agiriak (1997/11/24-28)*. Bilbao: Summer University of the Basque Country.

Campbell, L. R. (2011): “La investigación histórica de las lenguas aisladas, o ¿es raro el vasco?”. In: Lakarra, J. A. & Gorrochategui, J. & Urgell, B. (eds.). P. 23-40.

Campión, A. (1977 [1884]): *Gramática de los cuatro dialectos literarios de la lengua euskara* (2 vols.). Bilbao: La Gran Enciclopedia Vasca.

Cano, R. (ed.) (2008): *Historia de la lengua española*. Barcelona: Ariel.

Caro-Baroja, J. (1990 [1945]): *Materiales para una historia de la lengua vasca en su relación con la latina*. Donostia-San Sebastián: Txertoa.

Castaños-Garay, F. (1957): “El genitivo en vizcaíno antiguo”. In: *BAP*. Nr. 13. P. 60-69.

------- (1979): *La formación del plural en la lengua vasca*. Bilbao: Diputación Foral del Señorío de Vizcaya.

Cerrón-Palomino, R. (1987): *Lingüística quechua*. Cuzco: Centro de Estudios Andinos «Bartolomé de Las Casas».

Chamoreau, C. & Léglise, I. (eds.) (2012): *Dymamics of Contact-Induced Language Change*. Berlin, Boston: Walter de Gruyter.

Comrie, B. (1989 [1981a]): *Universales del lenguaje y tipología lingüística* (translated from the original version in English: *Language Universals and Linguistic Typology. Syntax and Morphology*). Madrid: Gredos.

------- (1981b): “Aspect and Voice: some reflections on perfect and passive”. In: Fedeschi, P. & Zaenen, A. (eds.). P. 65-78.

Corbett, G. G. (2000): *Number*. Cambridge University Press.

Corominas, J. (1960): “La toponymie hispanique prérromane et la survivance du basque jusqu’au bas Moyen Âge (Phénomènes de bilinguisme dans les Pyrénées Centrales)”. In: *Actes et Mémoires du VIe Congrès International de Sciences Onomastiques* (Munich 1958). Munich: Verlag der Bayerischen Akademie der Wissenschaften. P. 105-146.

------- (1965-1970): *Estudis de Toponimia Catalana. I-II*. Barcelona: Barcino.

------- (1989-1997): *Onomasticon Cataloniae. I-VII*. Barcelona: Curial.

------- & Pascual, J. A. (1980-91): *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos.

Corpus Arakatzailea (= Scrutinizing Corpus) [available at http://klasikoak.armiarma.com/corpus.htm, accessed on 2012-08-30].

Coyos, J.-B. (2007): *Xiberotarra zuberera. Le dialecte basque souletin*. Donostia-San Sebastián: Elkar.

------- (2012): “Euskararen eta gaskoieraren arteko harremanak: zubereraren adibide batzuk”. In: Igartua, I. (ed.). P. 151-164.

Creissels, D. & Mounole, C. (2011): “Animacy and spatial cases: Typological tendencies, and the case of Basque”. In: Kittilä, S. & Västi, K. & Ylikoski, J. (eds.). P. 157-182.

Croft, W. & Denning, K. & Kemmer, S. (eds.) (1990): *Studies in Typology and Diachrony for J. H. Greenberg*. Amsterdam-Philadelphia: John Benjamins.

De Bustos, J. J. (2008): “Las glosas emilianenses y silenses”. In: Cano, R. (ed.). P. 291-307.

Dixon, R. M. W. (1994): *Ergativity*. Cambridge University Press.

------- (2002): *Australian Languages*. Cambridge University Press.

Domene, J. F. (2005): *Lingüística y matemáticas: el proceso de formación de la estructura morfológica verbal de la lengua vasca*. Unpublished PhD presented at the University of the Basque Country.

Donegan, P. (1978): *On the Natural Phonology of Vowels*. Colombus: Ohio State University.

------- (1993): “Rhythm and vocalic drift in Munda and Mon-Khmer”. In: *Linguistics in the Tibeto-Burman Area*. Nr. 16. P. 1-43.

------- & Stampe, D. (1983): “Rhythm and the holistic organization of language structure”. In: Richardson, J. et alii (eds.). P. 337-353.

------- & ------- (2004): “Rhythm and the Synthetic Drift of Munda”. In: Singh, R. (ed.). P. 3-36.

Douglass, W. A. & Etulain, R. W. (eds.) (1977): *Anglo-American Contributions to Basque Studies: Essays in Honor of Jon Bilbao*. Reno: University of Nevada.

Dryer, M. S. (1988): “Object-Verb Order and Adjective-Noun Order: Dispelling a Myth”. In: *Lingua*. Nr. 74. P. 185-217.

------- & Haspelmath, M. (eds.) (2011): *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library [at http://wals.info/, accessed on 2012-08-22].

Duanmu, S. (1998): “Wordhood in Chinese”. In: Packard, J. L. (ed.). P. 135-196.

Durie, M. & Ross, M. (eds.) (1996): *The comparative method revisited. Regularity and irregularity in language change*. Oxford University Press.

Echave, B. de (1607): *Discursos de la antigüedad de la lengua cántabra Bascongada*. Mexico: Henrrico Martinez’s Press.

Echenique, M. T. (1997): *Estudios lingüísticos vasco-románicos*. Madrid: Ediciones Istmo.

------- (2008): “Léxico vasco y latino-románico en contacto”. In: *Oihenart*. Nr. 23. P. 61-75.

------- (2012): “Euskera y románico: su contacto ininterrumpido a través de la historia”. In: Igartua, I. (ed.). P. 85-92.

Egurtzegi, A. (2011): “Euskal metatesiak: abiaburua haien ikerketarako” (in press, to appear in *ASJU* 45).

Eguzkitza, A. (1995): “Jabetza-izenordainkiak bizkaieraz”. In: Gómez, R. & Lakarra, J. A. (eds.). P. 427-432.

Elordieta, G. (1997): “Accent, tone and intonation in Lekeitio Basque”, in Martínez-Gil, F. & Morales-Front, A. (eds.). P. 4-78.

------- (1998): “Intonation in a pitch-accent variety of Basque”. In: *ASJU*. Nr. 32. P. 511-569.

------- (2011a): “Euskal azentuaren bilakaera: hipotesiak eta proposamenak”. In: Sagarna, A. & Lakarra, J. & Salaberri-Zaratiegi, P. (eds.). P. 989-1014.

------- (2011b): “Euskal azentu eta intonazioari buruzko ikerketa: status quaestionis”. In: Lakarra, J. A. & Gorrochategui, J. & Urgell, B. (eds.). P. 389-428.

------- & Hualde, J. I. (2003): "Tonal and durational correlates of accent in context of downstep in Lekeitio Basque". In: *Journal of the International Phonetic Association*. Nr. 33. P. 195-209.

Epelde, I. & Zuazo, K. (eds.) (2011): *Euskalkien jatorria eta bilakaera*. Summer University of the Basque Country (in press, to appear in Supplements of *ASJU*).

Epps, P. (2006): "Growing a numeral system. The historical development of numerals in an Amazonian language family". In: *Diachronica*. Nr. 23. P. 259-288.

Erdozia, J. L. (2001): *Sakana erdialdeko euskara*. Pamplona-Iruñea: Government of Navarre.

Ernout, A. & Meillet, A. (2001 [1932]): *Dictionnaire étymologique de la langue latine*. Paris: Klincksieck.

Esnaola, I. (1999): *Gazte euskaldunen lagun arteko hizkera gaur eta hemen*. Pamplona-Iruñea: Udako Euskal Unibertsitatea.

Ettlinger, M. & Fleisher, N. & Park-Doob, M. (eds.) (2005): *Proceedings of the Thirtieth Annual Meeting of the Berkeley Linguistics Society (February 13-16, 2004)*. Berkeley: Berkeley Linguistics Society.

Etxebarria, J. M. (1991): *Zeberio haraneko euskararen azterketa etno-linguistikoa*. Euba, Zornotza: I.K.A.

Etxeberria, P. & Knörr, H. (eds.) (2005): *Nerekin yaio nun. Txillardegiri omenaldia* (*Iker* 17). Bilbao: Euskaltzaindia.

Etxeberria, U. (2010): "*-a*, *-ak* eta *-(r)ik* euskal hizkeretan zehar". In: Fernández, B. & Albizu, P. & Etxepare, R. (eds.). P. 65-84.

Etxepare, R. (2003): "Valency and argument structure in the Basque verb". In: *GB*. P. 363-426.

------- & Gómez, R. & Lakarra, J. A. (eds.) (2009): *Beñat Oiharzabali Gorazarre / Festschrift for Bernard Oyharçabal* [= *ASJU*. Nr. 43].

Euskaltzaindia (1987-2005): *Orotariko Euskal Hiztegia* (= *General Dictionary of Basque*) (16 vols.). Bilbao: Euskaltzaindia [also at: http://www.euskaltzaindia.net/oeh/, accessed on 2012-08-22].

------- (2002): *Euskal Gramatika Laburra: Perpaus Bakuna*. Bilbao: Euskaltzaindia.

Eys, W. J. van (1879): *Grammaire comparée des dialectes basques*. Paris: Maisonneuve.

Fant, G. (1968): "Analysis and synthesis of speech processes". In: Malberg, B. (ed.). P. 171-272.

Fedeschi, P. & Zaenen, A. (eds.) (1981): *Syntax and Semantics (Vol. 14). Tense and aspect*. New York, Londres, Toronto, Sydney, San Francisco: Academic Press.

Feng, S. (1998): "Prosodic structure and compound words in Classical Chinese". In: Packard, J. L. (ed.). P. 197-260.

Ferguson, Ch. A. & Hyman, L. & Ohala, J. J. (eds.) (1975): *Nasálfest: Papers from a symposium on nasals and nasalization*. Stanford University Press.

Fernández, B. & Albizu, P. & Etxepare, R. (eds.) (2010): *Euskara eta euskarak. Aldakortasun sintaktikoa aztergai* (Supplements of *ASJU* 52). Bilbao: University of the Basque Country.

------- & Laka, I. (eds.) (2006): *Andolin gogoan. Essays in honour of Professor Eguzkitza*. Bilbao: University of the Basque Country.

Ferrer i Jané, J. (2009). "El sistema de numerales ibérico: avances en su conocimiento". In: *Palaeohispanica*. Nr. 9. P. 451-479.

Fikkert, P. & Jacobs, H. (eds.) (2003): *Development in prosodic systems*. Berlin: Mouton de Gruyter.

Frajzyngier, Z. (2002): *Studies in Chadic Morphology and Syntax*. Paris-Brussels: Peeters.

Franco, J. & Landa, A. (eds.) (1999): *Grammatical analyses in Basque and Romance Linguistics*. Amsterdam: John Benjamins.

Frellesvig, B. & Whitman, J. (2008): "Evidence for seven vowels in proto-Japanese". In: Frellesvig, B. & Whitman, J. (eds.). P. 15-42.

------- & ------- (eds.) (2008): *Proto-Japanese. Issues and Prospects*. Amsterdam-Philadelphia: John Benjamins.

Gaminde, I. (1997): *Gatikako euskaraz*. Gatika: Gatikako Gogoz Euskara Taldea.

Gamkrelidze, T. V. & Ivanov, V. V. (1984): *Indo-European and the Indo-Europeans: A Reconstruction and Historical Typological Analysis of a Protolanguage and Proto-Culture* (2 vols.). Tbilisi State University.

Garrett, A. (2005): "The Evolution of Algic Verbal Stem Structure: New Evidence from Yurok". In: Ettlinger, M. & Fleisher, N. & Park-Doob, M. (eds.). P. 46-60.

Gavel, H. (1920): *Éléments de phonétique basque* [= *RIEV*. Nr. 12].

------- (1929): *Grammaire basque*. Bayonne: Imprimerie du Courrier.

Gipuzkoako Foru Aldundia (ed.) (2004): *Lazarragaren eskuizkribua*. Madrid: Edilan-Ars Libris.

Goenaga, P. (1996): "Subjunktiboaz eta aginteraz zenbait ohar: -EN eta -ELA atzizkien erabileraz". In: Altuna, P. (ed.). P. 109-132.

Gómez, R. (1991): "Erronkarierazko dotrina argitaragabe bat: edizioa eta azterketa". In: Lakarra, J. A. & Ruiz-Arzalluz, I. (eds.). P. 375-426.

------- (1994): "Euskal aditz morfologia eta hitzordena: VSO-tik SOV-ra?". In: Orpustan, J. B. (ed.). P. 93-114.

------- (1996): "La aportación de W. von Humboldt a la gramática vasca". In: *RIEV*. Nr. 41. P. 607-622.

------- (2005): "De re etymologica: vasc. *-(r)antz* 'hacia'". In: Etxeberria, P. & Knörr, H. (eds.). P. 273-280.

------- (2006): *XIX. mendeko euskal gramatikagintzari buruzko ikerketak*. PhD presented at the University of the Basque Country [available at http://artxiker.ccsd.cnrs.fr/artxibo-00105921, accessed on 2012-09-21].

------- (2008a): "Berriz Aizkibelen *Euskera* eskuizkribuaz". In: *ASJU*. Nr. 42. P. 333-380.

------- (2008b): "Euskal gramatikagintza zaharraren historia laburra: XVII.-XVIII. mendeak". In: Artiagoitia, X. & Lakarra, J. I. (eds.). P. 329-349.

------- & Sainz, K. (1995): "On the Origin of the Finite Forms of the Basque Verb". In: Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.). P. 235-278.

------- & Lakarra, J. A. (eds.) (1995): *Euskal Dialektologiako Kongresua* (Donostia, 1991ko irailak 2-6) (Supplements of *ASJU* 27). Donostia-San Sebastián.

Gómez-Pantoja, J. & Alfaro, E. (2001): "Indigenismo y Romanización en las tierras altas de Soria". In: Villar, F. & Fernández-Álvarez, M. P. (eds.). P. 168-187.

Good, J. (2005): "Reconstructing morpheme order in Bantu. The case of causativization and applicativization". In: *Diachronica*. Nr. 22. P. 3-57.

Gorrochategui, J. (1984): *Estudio sobre la onomástica indígena aquitana*. Bilbao: University of the Basque Country, University of Salamanca.

------- (1987): “Vasco-Celtica”. In: *ASJU*. Nr. 21. P. 951-960.
------- (1993): “La onomástica aquitana y su relación con la ibérica”. In: Untermann, J. & Villar, F. (eds.). P. 609-634.
------- (1995a): “The Basque Language and its Neighbors in Antiquity”. In: Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.). P. 31-63.
------- (1995b): “Los Pirineos entre Galia e Hispania: las lenguas”. In: *Veleia*. Nr. 12. P. 181-234.
------- (2002): “El área de Bilbao en la Antigüedad”. In: Arejita, A. *et alii* (eds.). P. 103-120.
------- (2004): “Las raíces lingüísticas de Navarra”. In: Jimeno, R. & López-Mugartza, J. L. (eds.). P. 105-122.
------- (2007-08): “Lenguas y genes: aplicaciones a la prehistoria de la lengua vasca”. In: *Veleia*. Nr. 24-25. P. 1185-1201.
------- (2009a): “Vasco antiguo: algunas cuestiones de geografía e historia lingüísticas”. In: *Palaeohispanica*. Nr. 9. P. 539-555.
------- (2009b): “Las lenguas de los Pirineos en los tiempos antiguos”. In: Santos-Yanguas, J. (ed.). P. 55-79.
------- (2011): “Las armas de la Filología: la cuestión del vasco antiguo y los hallazgos de Iruña-Veleia”. In: Lakarra, J. A. & Gorrochategui, J. & Urgell, B. (eds.). P. 41-70.
------- (2012): “Euskaratiko osagaiak gaztelaniaren lexikoan”. In: Igartua, I. (ed.). P. 103-150.
------- & Lakarra, J. A. (1996): “Nuevas aportaciones a la reconstrucción del protovasco”. In Villar, F. & Encarnaçao, J. D. (eds.). P. 101-145.
------- & ------- (2001): “Comparación lingüística, filología y reconstrucción del protovasco”. In: Villar, F. & Fernández-Álvarez, M. P. (eds.). P. 407-438.
------- & ------- (2013): “Why Basque cannot be, unfortunately, and Indo-European language” (to appear in *Journal of Indo-European Studies*).
Goyheneche, E. (2011): *Onomastique du nord du Pays Basque (XIéme - XVéme siècles)*. Bilbao: Euskaltzaindia.
Greenberg, J. (1963): “Some universals of grammar with particular reference to the order of meaningful elements”. In: Greenberg (ed.). P. 73-113.
------- (ed.) (1963): *Universals of Language*. Cambridge: MIT Press.
Grosu, A. & Thompson, S. A. (1977): “Constraints on the Distribution of NP Clauses”. In: *Language*. Nr. 53. P. 104-151.
Guiter, E. (1989): “Elementos de cronología fonética del vascuence”. In: *ASJU*. Nr. 23. P. 797-800.
Gussenhoven, C. (2004): “Northern Bizkaian Basque”. In: Gussenhoven, C. (ed.). P. 170-184.
------- (ed.) (2004): *The Phonology of Tone and Intonation*. Cambridge University Press.
------- & Riad, T. (eds.) (2007): *Tones and tunes*. Berlin: Mouton de Gruyter.
Haase, M. (1992): *Sprachkontakt und Sprachwandel im Baskenland: Die Einflüsse des Gascognischen und Französischen auf das Baskische*. Hamburg: Helmut Buske Verlag.
------- (1994): “Tense and Aspect in Basque”. In: Thieroff, R. & Ballweg, J. (eds.). P. 279-292.
Hamp, E. (1998): “Some draft principles for classification”. In: Salmons, J. C. & Joseph, B. D. (eds.). P. 13-16.
Harriet, M. de (1741): *Gramatica escuaraz eta francesez, composatua francez hitzcunça ikhasi nahi dutenen faboretan*. Bayonne: Fauvet Alarguna eta J. Fauvet.

Harris, A. C. (1990): “Kartvelian contacts with Indo-European”. In: Markey, T. L. & Greppin, J. A. C. (eds.). P. 67-100.
------- & Campbell, L. R. (1995): *Historical syntax in cross-linguistic perspective*. Cambridge University Press.
Haspelmath, M. (2008): “Frequency vs. iconicity in explaining grammatical asymmetries”. In: *Cognitive Linguistics*. Nr. 19(1). P. 1-33.
Hawkins, J. A. (1983): *Word Order Universals*. New York, London, etc.: Academic Press.
------- & Gilligan, G. (1988): “Prefixing and suffixing universals in relation to basic word order”. In: *Lingua*. Nr. 74. P. 219-259.
Hayes, B. & Kirchner, R. & Steriade, D. (eds.) (2004): *Phonetically Based Phonology*. Cambridge University Press.
Heath, J. (1977): “Remarks on Basque Verbal Morphology”. In: Douglass, W. A. & Etulain, R. W. (eds.). P. 193-201.
Heine, B. & Claudi, U. & Hünnemeyer, F. (1991): *Grammaticalization. A Conceptual Framework*. University of Chicago Press.
------- & Kuteva, T. (2002): *World Lexicon of Grammaticalization*. Cambridge University Press.
Hengeveld, K. & Rijkhoff, J. & Siewierska, A. (2004): “Parts-of-speech systems and word order”. In: *Journal of Linguistics*. Nr. 40. P. 527-570.
Hidalgo, B. (1994): *Hitzen ordena euskaraz*. PhD presented at the University of the Basque Country [at http://www.inguma.org/tesiak/Hidalgo_Eizagirre_1995.pdf, accessed on 2012-08-19].
------- (1995): “Ohar estatistiko garrantzitsuak euskararen hitz ordenaren inguru. Euskara, S.V.O.?”. In: *FLV*. Nr. 70. P. 401-420.
Hik Hasi (2011): “Hiru urteko haurren herenek D ereduan ikasiko dute Nafarroan”. In: *Hik Hasi*. Nr. 139 [available at http://www.hikhasi.com/artikulua/1687, accessed on 2011-03-17].
Holmer, N. M. (1970): “A historic-comparative analysis of the structure of the Basque language”. In: *FLV*. Nr. 2. P. 5-40.
Houis, M. (1970): “Reflexion sur une double correlation typologique”. In: *Journal of West African Languages*. Nr. 7. P. 59-68.
Hoz, J. de (1993): “La lengua y la escritura ibéricas y las lenguas de los íberos”. In: Untermann, J. & Villar, F. (eds.). P. 635-666.
------- (2001): “Hacia una tipología del ibérico”. In: Villar, F. & Fernández-Álvarez, M. P. (eds.). P. 335-362.
Hualde, J. I. (1989): “Acentos vizcaínos”. In: *ASJU*. Nr. 23. P. 275-325.
------- (1991): *Basque Phonology*. Londres, New York: Routledge.
------- (1993a): “On the Historical Origin of Basque Accentuation”. In: *Diachronica*. Nr. 10. P. 13-50.
------- (1993b): “Topics in Souletin Phonology”. In: Hualde, J. I. & Ortiz de Urbina, J. (eds.). P. 289-327.
------- (1993c): “Phonologically Unmotivated Changes in Language Contact: Spanish borrowings in Basque”. In: *Folia Linguistica*. Nr. 27. P. 1-25.
------- (1995): “Reconstructing the Ancient Basque Accentual System: Evidence and hypotheses”. In: Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.). P. 171-188.

------- (1997a): “Aitzineuskararen leherkariak”. In: *ASJU*. Nr. 31(2). P. 411-424.
------- (1997b): *Euskararen azentuerak* (Supplements of *ASJU* 42). Bilbao: University of the Basque Country.
------- (1999): “Pre-Basque plosives”. In: Franco, J. & Landa, A. (eds.). P. 77-104.
------- (2003a): “From phrase-final to post-initial accent in Western Basque”. In: Fikkert, P. & Jacobs, H. (eds.). P. 249-281.
------- (2003b): “Segmental phonology”. In: *GB*. P. 15-65.
------- (2003c): “Accentuation”. In: *GB*. P. 65-72.
------- (2006): “Laringalak eta euskal azentuaren bilakaera”. In: Fernández, B. & Laka, I. (eds.). P. 409-511.
------- (2007): “Historical convergence and divergence in Basque accentuation”. In: Gussenhoven, C. & Riad, T. (eds.). P. 291-322.
------- & Elordieta, G. & Elordieta, A. (1994): *The Basque Dialect of Lekeitio* (Supplements of *ASJU* 34). Bilbao.
------- & Lujanbio, O. (2008): “Goizuetako azentuera”. In: Artiagoitia, X. & Lakarra, J. A. (eds.). P. 377-394.
------- & ------- & Torreira, F. (2008): “Lexical tone and stress in Goizueta Basque”. In: *Journal of the International Phonetic Association*. Nr. 38. P. 1-24.
------- & Lakarra, J. A. & Trask, R. L. (eds.) (1995): *Towards a History of the Basque Language* (Amsterdam Studies in the Theory and History of Linguistic Science, Series IV: Current Issues in Linguistic Theory). Amsterdam: John Benjamins.
------- & Ortiz de Urbina, J. (eds.) (1993): *Generative Studies in Basque linguistics*. Amsterdam: John Benjamins.
------- & ------- (eds.) (2003): *A Grammar of Basque*. Berlin: Mouton de Gruyter.
Hyman, L. (2004): “How to become a “Kwa” verb”. In: *Journal of West African Languages*. Nr. 30. P. 69-88.
------- (2007): “Niger-Congo verb extensions: overview and discussion”. In: Payne, D. L. & Peña, J. (eds.). P. 149-163.
------- (2008): “Directional asymmetries in the morphology and phonology of words, with special reference to Bantu”. In: *Linguistics*. Nr. 46(2). P. 309-350.
Ibarra, O. (1995): *Ultzamako hizkera. Inguruko euskalkiekiko harremanak*. Pamplona-Iruñea: Government of Navarre.
------- (2010): *Nafar gazteen ahozko diskurtsoa*. Pamplona-Iruñea: Public University of Navarre.
Igartua, I. (2001): “La aspiración en vasco: ensayo tipológico y diacrónico”. In: *ASJU*. Nr. 35. P. 185-213.
------- (2006): “Del origen de la aspiración como elemento morfonológico en vasco”. In: Lakarra, J. A. & Hualde, J. I. (eds.). P. 519-530.
------- (2008): “La aspiración de origen nasal en la evolución fonética del euskera: un caso de rhinoglottophilia”. In: *ASJU*. Nr. 42. P. 171-189.
------- (2011): “Historia abreviada de la aspiración en las lenguas circumpirenaicas”. In: Sagarna, A. & Lakarra, J. A. & Salaberri-Zaratiegi, P. (eds.). P. 895-918.
------- (2012): “Euskara eta haren auzoak: aitzinsolas modura”. In: Igartua, I. (ed.). P. 7-16.
------- (ed.) (2012): *Euskara eta inguruko hizkuntzak historian zehar*. Vitoria-Gasteiz: Eusko Jaurlaritza - Gobierno Vasco.

------- & Santazilia, E. (2013): "Instrumentalaz gogoratuz, instrumentuaz haratago: kasuaren izaera tipologian eta euskararen morfosintaxian". Unpublished manuscript (to be published in the *3rd Conference of the Luis Michelena Chair*).

Igartua, M. (1986): "Leizarragaren *-ra* eta *-rat*en erabilerari buruz". In: *ASJU*. Nr. 20. P. 897-902.

Iggesen, O. A. (2011): "Number of Cases". In: Dryer, M. S. & Haspelmath, M. (eds.). Chapter 49.

Intxausti, J. (1990): *Euskara, la lengua de los vascos*. Vitoria-Gasteiz: Basque Government.

Iriarte, A. (1998): "La necrópolis de San Pelayo (Alegría-Dulantzi, Álava) y la cuestión de la fecha de inicio de las necrópolis de tipo merovingio en Álava". In: *Cuadernos de arqueología de la Universidad de Navarra*. Nr. 6. P. 139-163.

Irigoyen, A. (1987 [1976]): "Consideraciones históricas en torno al genitivo posesivo vasco" [in *DRFLU. II*. P. 177-179].

------- (1987 [1978]): "Erakusleez" [in *DRFLU. II*. P. 108].

------- (1981): "*Haur ola zirola*. Elementos deícticos en lengua vasca". In: *Bascologists International Meetings* (Lejona 1980) (*Iker* 1). Pamplona-Iruñea: Euskaltzaindia. P. 365-402.

------- (1985a): "Las lenguas de los vizcaínos: antroponimia y toponimia medievales". Bilbao: University of Deusto.

------- (1987 [1985b]): "Sobre la no determinación del sistema nominal en lengua vasca" [in *DRFLU. II*. 39-57].

------- (selected writings) (1985-1995): *De Re Philologica Linguae Uasconicae. I-V*. Bilbao: University of Deusto [= *DRFLU. I-V*].

------- (1996): "Gnomica Biscaiensis. *Refranes y Sentencias* de 1596". In: *Euskera*. Nr. 41. P. 689-718.

Iverson, G. & Salmons, J. (1992): "The phonology of the Proto-Indo-European root structure constraints". In: *Lingua*. Nr. 87. P. 293-320.

Jacobsen, W. H. (1975): "Historical Implications of the Western Basque Tonal Accent". Unpublished manuscript.

------- (1977): "The Basque locative suffix". In: Douglass, W. A. *et alii* (eds.). P. 163-168.

Janhunen, J. (2007): "The primary laringeal in Uralic and beyond". In: *Memoires de la Société Finno-Ougrienne*. Nr. 254. P. 203-227.

------- (2009): "Proto-Uralic: what, where, and when?". In: *The Quasquicentennial of the Finno-Ugrian Society. Suomalais-Ugrilaisen Seuran Toimituksia = Mémoires de la Société Finno-Ougrienne*. Nr. 258. P. 57-78.

Jauregi, O. (2007): *Silaba euskaraz: egitura eta historia*. PhD presented at the University of the Basque Country.

Jimeno, R. & López-Mugartza, J. C. (eds.) (2004): *Vascuence y romance: Ebro-Garona, un espacio de comunicación*. Pamplona-Iruñea: Government of Navarre.

Jones, Ch. (ed.) (1993): *Historical Linguistics: Problems and Perspectives*. London: Longman.

Joseph, B. D. & Zwicky, A. M. (eds.) (1990): *When Verbs Collide: Papers from the (1990) Ohio State Mini-Conference on Serial Verb Constructions*. Ohio State University.

------- & Janda, R. (eds.) (2003): *The Handbook of Historical Linguistics*. Oxford: Blackwell.

Kerejeta, M. J. (1991): "Notas sobre las poesías premiadas en Pamplona en 1609 y 1610". In: Lakarra, J. A. & Ruiz-Arzalluz (eds.). P. 161-183.

Kintana, X. (2005): “Lazarraga poeta zahar gaztearen berri onak”. In: *Euskera*. Nr. 50. P. 129-159.
Kitson, P. R. (1996): “British and European River Names”. In: *Transactions of the Philological Society*. Nr. 94. P. 73-118.
Kittilä, S. & Västi, K. & Ylikoski, J. (eds.) (2011): *Case, Animacy and Semantic Roles* (Typological Studies in Language 99). Amsterdam/Philadelphia: John Benjamins.
Knörr, H. (1985): “Para una delimitación etno-lingüística de la Álava antigua: ensayo de cartografía a partir de pruebas toponímicas”. In: *La formación de Álava. I*. Vitoria-Gasteiz: Diputación Foral de Álava. P. 489-541.
------- (1995): “La huella del latín en la lengua vasca”. In: Valcárcel, V. (ed.). P. 214-225.
------- (1996): “Joannes Leizarraga”. In: Aulestia, G. (ed.). P. 113-138.
------- & Zuazo, K. (eds.) (1998): *Arabako euskararen lekukoak. Ikerketak eta testuak*. Bilbao: Basque Parlament.
Lacombe, G. (1952 [1924]): “Langue basque”. In Meillet, A. & Cohen, M. (eds.). P. 255-270.
Ladefoged, P. & Maddieson, I. (1996): *The sounds of the world’s languages*. Oxford: Blackwell.
Lafitte, P. (1979 [1944]): *Grammaire basque (Navarro-labourdin littéraire)*. Donostia-San Sebastián: Elkar.
Lafon, R. (1999 [1937]): “Tendance à la palatalisation de la sonante *u* dans les parlers basques du nord-est” [in *Vasconiana*. P. 85-93].
------- (1980 [1944]): *Le systeme du verbe basque au XVI^e siècle* (2 vols.). Donostia-San Sebastián: Elkar.
------- (1999 [1948]): “Sur les suffixes casuels *-ti* et *-tik*” [in *Vasconiana*. P. 199-207].
------- (1999 [1950]): “Remarques sur la racine en basque” [in *Vasconiana*. P. 287-292].
------- (1999 [1951]): “La langue de Bernard Dechepare” [in *Vasconiana*. P. 729-758].
------- (1951-52): “Concordances morphologiques entre le basque et les langues caucasiques”. In: *Word*. Nr. 7. P. 227-244 ; and Nr. 8. P. 80-94.
------- (1999 [1954a]): “Le nombre dans la déclinaison basque” [in *Vasconiana*. P. 209-217].
------- (1999 [1954b]): “Comportement syntaxique, structure et diathese du verbe basque” [in *Vasconiana*. P. 529-555].
------- (1999 [1955]): “Sur la place de l’aezcoan, du salazarais, et du roncalais dans la classification des dialectes basques” [in *Vasconiana*. P. 707-726].
------- (1956a): “Pour l’étude de la langue aquitaine”. In: *Actes du deuxième Congrès International d’études Pyrénéennes (1954, Luchon-Pau) Tome Section VII*. Toulouse. P. 53-63.
------- (1999 [1956b]): “Les formes verbales de prétérit à préfixe *l-* dans les textes du XVIe siècle” [in *Vasconiana*. P. 459-473].
------- (1999 [1958]): “Contribution à l’étude phonologique du parler basque de Larrau (Haute-Soule)” [in *Vasconiana*. P. 113-133].
------- (1999 [1961]): “Sur les formes verbales basques qui contiennent un indice datif” [in *Vasconiana*. P. 411-430].
------- (1999 [1962a]): “Contact de langues et apparition d’une nouvelle voyelle: *u* et *ü* en basque” [in *Vasconiana*. P. 135-143].
------- (1999 [1962b]): “Sur la voyelle *ü* en basque” [in *Vasconiana*. P. 95-111].
------- (1999 [1965]): “Les deux génitifs en basque” [in *Vasconiana*. P. 173-197].
------- (1999 [1970]): “Structure de la déclinaison basque” [in *Vasconiana*. P. 167-171].

------- (collected writings edited by Haritschelhar, J. & Charritton, P.) (1999): *Vasconiana* (*Iker* 11). Bilbao: Euskaltzaindia.
Laka, I. (1987): "Manuel Arriandiaga: Berrazterketarako oinarriak (Hiperbizkaieraren historiaz IV)". In: *ASJU*. Nr. 21. P. 727-753.
------- (1988): "Configurational Heads in Inflectional Morphology: the structure of the inflected forms in Basque". In: *ASJU*. Nr. 22. P. 343-365.
------- (1993): "The Structure of Inflection: A case study in Xº syntax". In: Hualde, J. I. & Ortiz de Urbina, J. (eds.). P. 21-70.
Lakarra, J. A. (1983): "Oharrak zenbait arkaismoz". In: *ASJU*. Nr. 17. P. 41-68.
------- (1984a): "Bizkaiera zaharreko ablatiboaz". In: *ASJU*. Nr. 18(1). P. 163-193.
------- (1984b): "Bertso Bizkaitarrak (1688)". In: *ASJU*. Nr. 18(2). P. 89-184.
------- (1985): "Literatur gipuzkerarantz: Larramendiren Azkoitiko Sermoia (1737)". In: *ASJU*. Nr. 19. 235-281.
------- (1986): "Bizkaiera zaharra euskalkien artean". In: *ASJU*. Nr. 20. P. 639-682.
------- (1995): "Reconstructing the Pre-Proto-Basque Root". In: Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.). P. 189-206.
------- (1996a): *Refranes y Sentencias. Ikerketak eta edizioa*. Bilbao: Euskaltzaindia.
------- (1996b): "Sobre el europeo antiguo y la reconstrucción del protovasco". In: *ASJU*. Nr. 30. P. 1-70.
------- (1997): "Euskararen historia eta filologia: arazo zahar, bide berri". In: *ASJU*. Nr. 31. P. 447-535.
------- (1998a): "Hizkuntzalaritza konparatua eta aitzineuskararen erroa". In: *Uztaro*. Nr. 25. P. 47-110.
------- (1998b): "Gure izterlehengusuek eta guk erro bera? Gogoetak erroaz aitzinkartvelikoz eta aitzineuskaraz". In Túrrez, I. & Arejita, A. & Isasi, C. (eds.). P. 125-150.
------- (2002): "Etymologiae (proto)uasconicae LXV". In: Artiagoitia, X. & Goenaga, P. & Lakarra, J. A. (eds.). P. 425-442.
------- (2003): "Hacia un nuevo paradigma etimológico vasco: forma canónica, filología y reconstrucción". In: *ASJU*. Nr. 37. P. 261-392.
------- (2004a): "Juan Perez Lazarragakoaren eskuizkribua (XVI. mendea). Lehen hurbilketa". In: Gipuzkoako Foru Aldundia (ed.). P. V-XXV.
------- (2004b): "Bisílabos (¿proto?)vascos". Unpublished manuscript. University of the Basque Country.
------- (2005): "Prolegómenos a la reconstrucción de segundo grado y al análisis del cambio tipológico en (proto)vasco". In: *Palaeohispanica*. Nr. 5 (= *Acta Palaeohispanica IX*). P. 407-459.
------- (2006a). "Notas sobre iniciales, cambio tipológico y prehistoria del verbo vasco". In: Lakarra, J. A. & Hualde, J. I. (eds.). P. 561-622.
------- (2006b): "Protovasco, munda y otros: reconstrucción interna y tipología holística diacrónica". In: *Oihenart*. Nr. 21. P. 229-322.
------- (2006c): "*Jaun* eta *jabe, jaio* eta *herio, jin* eta *joan*… etimologiaz eta aditz zaharraz". In: Fernández, B. & Laka, I. (eds.). P. 575-611.
------- (2008a): "Aitzineuskararen gramatikarantz (malkar eta osinetan zehar)". In: Artiagoitia, X. & Lakarra, J. A. (eds.). P. 451-490.

------- (2008b): "*Vida con/y libertad*: sobre una coordinación arcaica y la autenticidad de «Urthubiako alhaba»". In: *ASJU*. Nr. 42. P. 83-100.
------- (2008c): "Forma canónica, etimología y reconstrucción en el campo vasco". In: *ASJU* Nr. 37. P. 261-391.
------- (2009a) "Temas para un prólogo: forma canónica, tipología holística diacrónica y reconstrucción del protovasco". In: *Oihenart*. Nr. 23. P. 277-347.
------- (2009b): "Forma canónica y cambios en la forma canónica en la prehistoria de la lengua vasca". In: *Palaeohispanica*. Nr. 9. P. 557-609.
------- (2009c): "Adabakiak /h/-aren balio etimologikoaz". In: Etxepare, R. & Gómez, R. & Lakarra, J. A. (eds.). P. 565-596.
------- (2009d): "$*h_3 > h_1$, $*h_2 > h_1$ eta horiei datxezkien zenbait fenomenoz". In: *Lapurdum*. Nr. 13. P. 211-231.
------- (2010a): "Aitzineuskara berreraikiaz: zergatik ezkerra". In: *Euskera*. Nr. 54. P. 17-98.
------- (2010b): "Haches, diptongos y otros detalles de alguna importancia: notas sobre numerales (proto)vascos y comparación vasco-ibérica (con un apéndice sobre *hiri* y otro sobre *bat-bi*)". In: *Veleia*. Nr. 27. P. 191-238.
------- (2010c): "Más sobre dentales iniciales". Unpublished manuscript. University of the Basque Country.
------- (2011a): "Erro monosilabikoaren teoria eta aitzineuskararen berreraiketa: zenbait alderdi eta ondorio". In: *FLV*. Nr. 113. P. 5-114.
------- (2011b): "Aitzineuskararen berreraiketaz: egindakoaz eta eginkizunez". In: Sagarna, A. & Lakarra, J. A. & Salaberri-Zaratiegi, P. (eds.). P. 617-694.
------- (2011c): "/h/ etimologiko gehiago: 5., 6. eta 7. /h/". Unpublished manuscript. University of the Basque Country.
------- (2011d): "Haches etimológicas vascas y testimonio gascón". Unpublished manuscript. University of the Basque Country.
------- (2011e): "Teoría de la raíz monosilábica y reconstrucción del protovasco: algunos aspectos y consecuencias". In: Lakarra, J. A. & Gorrochategui, J. & Urgell, B. (eds.). P. 651-699.
------- (2011f): "Gogoetak euskal dialektologia diakronikoaz: Euskara Batu Zaharra berreraiki beharraz eta haren banaketaren ikerketaz" (to be published in Epelde, I. & Zuazo, K., eds.).
------- (2012a): "On Ancient European and the reconstruction of Proto-Basque". In: Udolph, J. (ed.) (in press).
------- (2012b): "Retrotrayendo gramaticalizaciones: Ley de Trask, ablativos y adjetivos/verbos en *-ti/-di*, comparativos en *-so* (+ *oso*), completivos en *-to*/*-do* (+ *oro*), causativos y otros compañeros de viaje". Unpublished manuscript. University of the Basque Country.
------- (2012c): "Mailegaketa eta berreraiketa euskararen historiaurrearen ikerketan". In: Igartua, I. (ed.). P. 17-74.
------- (2013): "Aitzineuskararen berreraiketa sakonagorantz: forma kanonikoa, tipologia holistikoa, kronologia eta gramatikalizazioa". Unpublished manuscript, to be published in the *3rd Conference of the Luis Michelena Chair*.
------- & Biguri, K. & Urgell, B. (eds.) (1983): *Euskal Baladak. Antologia eta azterketa*. Donostia-San Sebastián: Hordago.

------- & Gorrochategui, J. & Urgell, B. (eds.) (2011): *2nd Conference of the Luis Michelena Chair (Vitoria-Gasteiz, October 2007)*. Vitoria-Gasteiz: University of the Basque Country.

------- & Hualde, J. I. (eds.) (2006): *Studies in Basque and Historical Linguistics in Memory of Larry Trask* [= *ASJU*. Nr. 40].

------- & Ruiz-Arzalluz, I. (eds.) (1991): *Memoriæ L. Mitxelena magistri sacrum* (Supplements of *ASJU* 14). Donostia-San sebastián.

------- & ------- (eds.) (2011): *Luis Michelena. Obras Completas* (15 vols.) (Supplements of *ASJU* 54-68). University of the Basque Country.

Larramendi, M. (1984 [1729]): *El impossible vencido. Arte de la lengua bascongada*. Bilbao: Editorial Amigos del Libro Vasco.

Larrasquet, J. (1932): "Phonétique du basque de Larrajá (quartier de Barcus)". In: *RIEV*. Nr. 23. P. 153-191.

Lehmann, W. P. (1973): "A Structural Principle of Language and its Implications". In: *Language*. Nr. 49. P. 47-66.

Li, Ch. N. & Thompson, S. A. (1976): "Subject and Topic: a New Typology of Language". In: Li, Ch. N. (ed.). P. 457-489.

------- (ed.) (1976): *Subject and Topic*. New York, San Francisco, London: Academic Press.

Líbano-Zumalacárregui, A. (1995-2000): *Toponimia Medieval en el País Vasco* (4 vols.). Bilbao: Euskaltzaindia.

Lichtenberk, F. (1985): "Syntactic-category change in Oceanic languages". In: *Oceanic Linguistics*. Nr. 24. P. 1-84.

López-Mugartza, J. C. (2008): *Erronkari eta Ansoko toponimiaz*. Pamplona-Iruñea: Government of Navarre.

Lord, C. (1993): *Historical Change in Serial Verb Constructions* (Typological Studies in Language 26). Amsterdam: John Benjamins.

Luchaire, D. J. A. (1877): "Les origines linguistiques de l'Aquitaine". In: *Bulletin de la Société des Sciences, Lettres et Arts de Pau* (IIe Série). Nr. 6. P. 349-423.

------- (1879): *Études sur les idiomes pyrénéens de la région française*. Paris: Maisonneuve et Cia.

Luraghi, S. & Huumo, T. (eds.) (2012): *Partitives*. Berlin-New York: Mouton de Gruyter.

Malberg, B. (ed.) (1968): *Manual of Phonetics*. Amsterdam: North Holland.

Malchukov, A. & Spencer, A. (eds.) (2009): *The Oxford Handbook of Case*. Oxford University Press.

Mallory, J. P. (1989): *In Search of the Indo-Europeans. Language, Archaeology and Myth*. London: Thames & Hudson.

Manning, C. D. (1996): *Ergativity. Argument structure and grammatical relations*. Stanford: CSLI Publications.

Manterola, J. (2006): "*-a* euskal artikulu definituaren gainean zenbait ohar". In: Lakarra, J. A. & Hualde, J. I. (eds.). P. 651-676.

------- (2008): "Euskarazko artikuluak Erdi Aroko agiri bilduma batean". In: *Oihenart*. Nr. 23. P. 371-379.

------- (2009a): "Euskal artikuluak euskalki eta autoretan". In: *Oihenart*. Nr. 24. P. 237-256.

------- (2009b): "Is Basque an Agglutinative Language? A Proposal for the Diachrony of Nominal Morphology". In: *TINTA*. Nr. 9. P. 88-108.

------- (2012): "The Basque articles *-a* and *bat* and recent contact theories". In: Chamoreau, C. & Léglise, I. (eds.). P. 231-264.
Markey, T. L. & Greppin, J. A. C. (eds.) (1990): *When Worlds Collide: Indo-Europeans and Pre-Indo-Europeans*. Ann Arbor: Karoma Publishers.
Martinet, A. (1974 [1950]): "La reconstrucción estructural: las oclusivas del vasco" [in *ECP*. P. 524-550].
------- (1974 [1955]): *Economía de los cambios fonéticos. Tratado de fonología diacrónica* (translated from the original version in French: *Économie des changements phonétiques*). Madrid: Gredos [= *ECP*].
------- (1981): "La phonologie synchronique et diachronique du basque". In: *Bascologists International Meetings* (Lejona 1980) (*Iker* 1). Pamplona-Iruñea: Euskaltzaindia. P. 59-74.
Martínez-Areta, M. (2004): "El acento proto-vasco". In: *ASJU*. Nr. 38. P. 135-206.
------- (2006): *El consonantismo proto-vasco*. PhD presented at the University of the Basque Country.
------- (2009a): "The Category of Number in Basque: I. Synchronic and historical aspects". In: *FLV*. Nr. 110. P. 63-98.
------- (2009b): "The Category of Number in Basque: II. Prehistorical and typological aspects" In: *FLV*. Nr. 111. P. 249-280.
------- (2011): "Euskararen hitz hurrenkeraren garaiketa baterako materialak". In: *Oihenart*. Nr. 26. P. 343-379.
------- (2012): "Nor-Nork sailaren sorreraz" (to be published in *ASJU* 46).
Martínez-Gil, F. & Morales-Front, A. (eds.) (1997): *Issues in the Phonology and Morphology of the Major Iberian Languages*. Washington DC: Georgetown University Press.
Martínez de Madina (2009-10): *Toponimia de Vitoria* (3 vols.). Bilbao: Euskaltzaindia.
Martínez-Sáenz de Jubera, M. & González-Perujo, J. M. (1998): "Onomástica vasca en La Rioja". In: *FLV*. Nr. 79. P. 475-496.
Masek, C. & Hendrick, R. A. & Miller, M. F. (eds.) (1981): *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society.
Matisoff, J. (1975): "Rhinoglottophilia: the mysterious connection between nasality and glottality". In: Ferguson, Ch. A. & Hyman, L. & Ohala, J. J. (eds.). P. 265-287.
------- (2003): *Handbook of Proto-Tibeto-Burman*. University of California Press.
Meillet, A. (1925): *La méthode comparative en linguistique historique*. Oslo: H. Aschehoug & Co. (W. Nygaard).
------- & Cohen, M. (1952 [1924]): *Les langues du monde*. Paris: Centre National de la Recherche Scientifique.
Melena, J. L. (ed.): *Symbolae L. Mitxelena* (2 vols.). Vitoria-Gasteiz: University of the Basque Country.
Merino-Urrutia, J. J. B. (1978 [1962]): *La lengua vasca en la Rioja y Burgos*. Logroño: Diputación Provincial de Logroño.
Michelena, L. (1988 [1950]): "De fonética vasca. La aspiración intervocálica" [in *SHLV*. P. 190-202].
------- (1988 [1951a]): "De fonética vasca. La distribución de las oclusivas aspiradas y no aspiradas" [in *SHLV*. P. 212-219].
------- (1988 [1951b]): "La sonorización de las oclusivas iniciales. A propósito de un importante artículo de André Martinet" [in *SHLV*. P. 203-211].

------- (1997 [1953a]): *Apellidos vascos*. Donostia-San Sebastián: Txertoa.
------- (1988 [1953b]): "Contribución al conocimiento del dialecto roncalés" [in *SHLV*. P. 246-272].
------- (1988 [1954a]): "La posición fonética del dialecto vasco del Roncal" [in *SHLV*. P. 273-297].
------- (1985 [1954b]): "De onomástica aquitana" [in *LH*. P. 409-445].
------- (1988 [1957a]): "Las antiguas consonantes vascas" [in *SHLV*. P. 166-189; also in English in Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.; 1995). P. 101-135].
------- (1988 [1957b]): "Basque et roman" [in *SHLV*. P. 106-115].
------- (1988 [1957/58]): "À propos de l'accent basque" [in *SHLV*. P. 220-239].
------- (1977 [1961]): *Fonética histórica vasca* (2nd edition) (Supplements of *ASJU* 4). Donostia-San Sebastián [= FHV].
------- (1985 [1961-62]): "Los nombres indígenas de la inscripción hispano-romana de Lerga (Navarra)" [in *LH*. P. 446-457].
------- (1990 [1963]): *Lenguas y protolenguas* (reed. as Supplements of *ASJU* 20). Donostia-San Sebastián.
------- (1990 [1964a]): *Textos arcaicos vascos* (reed. as Supplements of *ASJU* 11). Donostia-San Sebastián [= TAV].
------- (1988 [1964b]): *Sobre el pasado de la lengua vasca* [in *SHLV*. P. 1-73].
------- (1985 [1968]): "L'euskaro-caucasien" [in *LH*. P. 458-475].
------- (1987 [1969]): "Notas lingüísticas a «Colección diplomática de Irache»" [in *PT*. P. 87-140].
------- (1987 [1970]): "Nombre y verbo en la etimología vasca" [in *PT*. P. 283-309].
------- (1987 [1971a]): "Notas sobre las lenguas de la Navarra medieval" [in *PT*. P. 73-85].
------- (1987 [1971b]): "Toponimia, léxico y gramática" [in *PT*. P. 141-167].
------- (1988 [1971c]): "Descubrimiento y redescubrimiento en textos vascos" [in *SHLV*. P. 713-728].
------- (1988 [1972a]): "Nota marginal sobre la huella latina en la lengua vasca" [in *SHLV*. P. 121-137].
------- (1985 [1972b]): "Etimología y transformación" [in *LH*. P. 296-308].
------- (1987 [1972c]: "A Note on Old Labourdin Accentuation" [in *PT*. P. 235-244].
------- (1987 [1972d]: "Léxico vasco y etimología" [in *PT*. P. 337-348].
------- (1987 [1974]): "El elemento latino-románico en la lengua vasca" [in *PT*. P. 195-220].
------- (1985 [1976]): "Ibérico *-en*" [in *LH*. P. 379-387].
------- (1987 [1977]): "Notas sobre compuestos verbales vascos" [in *PT*. P. 311-335].
------- (1987 [1978]): "Miscelánea filológica vasca. I" [in *PT*. P. 363-385].
------- (1985 [1979a]): "La langue ibère" [in *LH*. P. 341-356].
------- (1987 [1979b]): "Miscelánea filológica vasca. IV" [in *PT*. P. 435-463].
------- (1979c): "Bibliografía". In: *BAP*. Nr. 35(3-4). P. 571-578.
------- (1987 [1981a]): "Lengua común y dialectos vascos" [in *PT*. P. 35-55].
------- (1987 [1981b]): "*Apodaka*" [in *PT*. P. 183-185].
------- (1987 [1982a]): "Sobre la historia de la lengua vasca" [in *PT*. P. 7-21].
------- (1987 [1982b]): "Sobre la lengua vasca en Álava durante la Edad Media" [in *PT*. P. 169-181].
------- (1988 [1984]): "Estratos en la toponimia alavesa" [in *SHLV*. P. 604-613].

------- (1988 [1985]): “Algunos nombres de Contrasta” [in *SHLV*. P. 614-619].
------- (1985): *Lengua e Historia*. Madrid: Paraninfo [= *LH*].
------- (1987): *Palabras y Textos*. Bilbao: University of the Basque Country [= *PT*].
------- (1988) (writings edited by Lakarra, J. A. *et alii*): *Sobre Historia de la Lengua Vasca* (2 vols.) (Supplements of *ASJU* 10). Donostia-San Sebastián [= *SHLV*].
------- (1992): “Fenómenos de convergencia en la historia de los dialectos vascos”. In: *FLV*. Nr. 61. P. 355-366.
Micoleta, R. (1995 [1653]): *Modo Breue de aprender la lengua Vizcayna*. Bilbao: AEK.
Mounole, C. (2006): “Quelques remarques à propos de l’histoire des périphrases basques”. In: Lakarra, J. A. & Hualde, J. I. (eds.). P. 723-738.
------- (2007): “Perifrasi zaharra mendebalde eta erdialdeko euskara zaharrean: azterketa kuantifikatiboa eta proposamen berria”. In: *ASJU*. Nr. 41(1). P. 67-138.
------- (2008): “Sintaxi diakronikoa eta aditz multzoaren garapena: inperfektibozko perifrasiaren sorreraz”. In: Artiagoitia, X. & Lakarra, J. A. (eds.). P. 585-604.
------- (2011): *Le verbe basque ancien: étude philologique et diachronique*. PhD presented at the University of the Basque Country / University Michel de Montaigne – Bordeaux III [available at https://www.educacion.gob.es/teseo/mostrarRef.do?ref=959808, accessed on 2012-11-01].
Mugarza, P. (2006): *Mallabiako euskara*. City Council of Mallabia.
Múgica, S. (1914-17): “El Obispado de Bayona en relación a los pueblos de Guipúzcoa adscritos a dicha diócesis”. In: *RIEV*. Nr. 8. P. 185-229.
Mujika, J. A. (2002): *Euskal ortografiaren hastapenak Iparraldeko literaturan*. Bilbao: University of the Basque Country.
------- (2011): “Inperatiboaren aditz-aldiak”. In: Salaburu, P. & Goenaga, P. & Sarasola, I. (eds.).
Narrog, H. (2009): “Varieties of Instrumental”. In: Malchukov, A. & Spencer, A. (eds.). P. 593-600.
------- (2010): “A diachronic dimension in maps of case functions”. In: *Linguistic Discovery*. Nr. 8: 1. P. 233-254 [available at http://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/1/xmlpage/1/article/352, accessed on 2012-10-14].
------- & Ito, S. 2007. “Reconstructing semantic maps. The Comitative-Instrumental Area”. In: *Sprachtypologie und Universalienforschung*. Nr. 60. P. 273-292.
Navarro-Tomás, T. (1923): “Observaciones fonéticas sobre el vascuence de Guernica”. In: *III Congreso de Estudios Vascos*. Donostia-San Sebastián: Eusko Ikaskuntza. P. 49-56
Newman, J. (ed.) (2002): *The Linguistics of Sitting, Standing and Lying* (Typological Studies in Language 51). Amsterdam/Philadelphia: John Benjamins.
Newman, P. & Newman, R. M. (eds.) (1977): *Papers in Chadic Linguistics*. Leiden: Afrika-Studiecentrum.
Núñez-Astrain, L. (2003): *El euskera arcaico. Extensión y parentescos*. Tafalla: Txalaparta.
Ohala, J. J. (1981): “The listener as a source of sound change”. In: Masek, C. & Hendrick, R. A. & Miller, M. F. (eds.). P. 178-203.
------- (1993): “The phonetics of sound change”. In: Jones, Ch. (ed.). P. 237-278.
Oihenart, A. (1992 [1638/1656]): *Notitia Utriusque Vasconiae*. Vitoria-Gasteiz: Basque Parlament.

Oñederra, L. (2009a): "Diacronía y motivación fonética". In: Sánchez Miret, F. (ed.). P. 47-67.
------- (2009b): "Ohar batzuk bokal batez: Zubereraren [y]". In: *ASJU*. Nr. 43. P. 663-678.
Orduña, E. (2005). "Sobre algunos posibles numerales en textos ibéricos". In: *Palaeohispanica*. Nr. 5. P. 491-506.
------- (2010): "En torno al lexema ibérico *eki-* y sus variantes". In: *Palaeohispanica*. Nr. 10. P. 319-334.
------- (2011): "Los numerales ibéricos y el protovasco". In: *Veleia*. Nr. 28. P. 125-139.
Oregi, J. (1974): "Euskal aditzaz zenbait gogoeta". *FLV*. Nr. 17. P. 265-283.
Ormaetxea, T. (2002): *Aramaioko euskara*. City Council of Aramaio.
Orpustan, J. B. (1991): *Toponymie basque. Noms de pays, communes, hameaux et quartiers historiques de Labourd, Basse-Navarre et Soule*. University of Bordeaux III.
------- (1999): *La langue basque au moyen age (IXe-XVe siècles)*. Baigorri: Izpegi.
------- (2006): *Nouvelle toponymie basque. Noms de pays, vallées, communes et hameaux historiques de Labourd, Basse-Navarre et Soule*. University of Bordeaux III.
------- (ed.) (1994): *Le basque parmi les autres langues: influences, comparaisons*. Bordeaux, Bayonne: University of Bordeaux III & CNRS.
Ortega, A. & Amorrortu, E. & Goirigolzarri, J. & Urla, J., & Uranga, B. (2013) (forthcoming): "Basque Linguistic Identity among New Speakers of Basque" (to be published in *International Journal of the Sociology of Language*).
------- & ------- (2010): "The attitudes of non-minority language speakers to minority language revitalization: what are non-Basque speakers prepared to do for Basque?". *Reversing Language Shift: How to Re-awaken a Language Tradition. Proceedings of the Foundation of Endangered Languages Annual Conference* (Carmathen, 13-15 September 2010). Bath: Foundation for Endangered Languages. P. 52-59.
Ortiz de Urbina, J. (1989): *Parameters in the grammar of Basque*. Dordrecht: Foris.
------- (2002): "Focus of Correction and Remnant Movement in Basque". In: Artiagoitia, X. & Goenaga, P. & Lakarra, J. A. (eds.). P. 510-524.
------- (2003): "Periphrastic constructions". In: *GB*. P. 284-300.
Osa-Unamuno, E. (1990): *Euskararen hitzordena komunikazio zereginaren arabera*. University of the Basque Country.
Oyharçabal, B. (1998): "R. L. Trask. The History of Basque". In: *Lapurdum*. Nr. 3. P. 317-333.
------- (2001): "Zenbait gogoeta euskarak letra hizkuntza gisa izan duen bilakaeraz (XVII-XVIII. mendeak)". In: *Litterae Vasconicae*. Nr. 8. P. 9-46.
------- (2003): "Tense, aspect and mood". In: Hualde, J. I. & Ortiz de Urbina, J. (eds.). P. 249-284.
Packard, J. L. (ed.) (1997): *New approaches to Chinese Word formation: morphology, phonology and the lexicon in Modern and Ancient Chinese*. Berlin: Mouton de Gruyter.
Padilla-Moyano, M. (ed.) (2011): *Kadet eta Bettiriño (ikerketa eta edizio kritikoa)*. Bilbao: Euskaltzaindia.
------- (2012): "Euskararen laguntzaile ahantziaz: **iron* aditzaren historia". Unpublished manuscript. University of the Basque Country.
Pagola, R. M. (1991): *Dialektologiaren atarian. Euskal dialektologiaren historiara hurbilketa*. Bilbao: Gero.

------- *et alii* (eds.) (1992-1999): *Bonaparte Ondareko Eskuizkribuak* (26 vols.). Bilbao: University of Deusto.

Palancar, E. L. (2002): *The Origin of Agent Markers* (Studia Typologica 5). Berlin: Akademie Verlag.

Palay, S. (1991): *Dictionnaire du béarnais et du gascon modernes* (3rd edition). Paris: CNRS Éditions.

Pawley, A. (2006): "Where have all the verbs gone? Remarks on the organization of languages with small, closed verb classes". Paper presented at *11th Binnenial Rice University Linguistics Symposium* (16-18 March 2006) [available at http://www.ruf.rice.edu/~lingsymp/Pawley_paper.pdf, accessed on 2012-08-29].

Payne, D. L. & Peña, J. (eds.) (2007): *Selected Proceedings of the 37th Annual Conference on African Linguistics*. Somerville, MA: Cascadilla Proceedings Project.

Penny, R. (2006): *Gramática histórica del español*. Barcelona: Ariel.

Pérez-Bilbao, A. (1991): *Bermeoko herri hizkera*. Bilbao: Igarri.

Peterson, D. (2007): "Toponimia vasca medieval: novedades del Becerro Galicano de San Millán de la Cogolla". In: *ASJU*. Nr. 41. P. 289-321.

------- (2009): *Frontera y lengua en el Alto Ebro, siglos VIII-XI. Las consecuencias e implicaciones de la invasión musulmana*. Logroño: IER.

Pikabea, J. (1993): *Lapurtera idatzia (XVII-XIX). Bilakaera baten urratsak*. Donostia-San Sebastián: University of the Basque Country.

Plank, F. (ed.) (1979). *Ergativity. Towards a Theory of Grammatical Relations*. London, New York, Toronto, Sydney, San Francisco: Academic Press.

Polák, V. (1963): "Les origines de la différenciation dialectale en Albanie". In: *Orbis*. Nr. 12. P. 381-392.

Pop, S. (1950): *La dialectologie: aperçu historique et méthodes d'enquêtes linguistiques* (2 vols.). Louvain: J. Duculot.

Post, M. W. (2006): "Compounding and the structure of the Tani lexicon". In: *Linguistics in the Tibeto-Burman Area*. Nr. 29. P. 40-60.

------- (2011): "Prosody and Typological Drift in Austroasiatic and Tibeto-Burman: against Indosphere and Sinosphere". In: *Austroasiatic Studies: papers from ICAAL* (*Mon-Khmer Studies Special Issue*. Nr. 3). Canberra: Pacific Linguistics. P. 198-229.

Poza, A. de (1587): *De la antigua lengua de las Españas*. Bilbao: Mathias Mares' Press.

Rebuschi, G. (1986): "Diathese et (non-)configurationnalité: L'exemple du basque" [in *ELB*. P. 73-99].

------- (1995): "Weak and Strong Genitive Pronouns in Northern Basque: A diachronic perspective". In: Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.). P. 313-356.

------- (selected writings) (1997): *Essais de linguistique basque* (Supplements of *ASJU* 35). Bilbao [= *ELB*].

Reguero-Ugarte, U. (2012): *Erdi Aroko euskara: dialektalizazioaren hastapenerantz*. Master Thesis presented at the University of the Basque Country.

Rey, A. (dir.) (2006): *Dictionnaire historique de la langue française* (3 vols.). Paris: Dictionnaires Le Robert.

Richardson, J. *et alii* (eds.) (1983): *Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax*. University of Chicago.

Rijk, R. de (1998 [1969]): "Is Basque an S.O.V. language?" [in *DLV*. P. 13-37].

------- (1998 [1972a]): “Studies in Basque Syntax: Relative Clauses” [in *DLV*. P. 71-149].
------- (1998 [1972b]): “Partitive Assignment in Basque” [in *DLV*. P. 151-182].
------- (1998 [1980]): “Erlatiboak idazle zaharrengan” [in *DLV*. P. 203-210].
------- (1998 [1981]): “Euskal morfologiaren zenbait gorabehera” [in *DLV*. P. 211-224].
------- (1998 [1985]): “Un verbe méconnu” [in *DLV*. P. 225-241].
------- (1995a): “«Nunc» in Old Basque”. In: Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.). P. 295-311.
------- (1998 [1995b]): “Basque manner adverbs and their genesis” [in *DLV*. P. 391-419].
------- (1998 [1996]): “On the Origin of the Partitive Determiner” [in *DLV*. P. 435-447].
------- (1998). *De Lingua Vasconum: Selected Writings* (Supplements of *ASJU* 43). Bilbao [= *DLV*].
------- (2002): “L’antipassif basque et l’hypothèse de Levin”. In: *Lapurdum*. Nr. 7. P. 295-312.
------- (2006): “On the so-called verbal noun in Basque”. In: Lakarra, J. A. & Hualde, J. I. (eds.). P. 865-870.
------- (2008): *Standard Basque. A Progressive Grammar*. Cambridge, London: MIT Press.
Ringe, D. (2003): “Internal reconstruction”. In: Joseph, B. & Janda, R. (eds.). P. 244-261.
Rollo, W. (2006 [1925]): *The Basque Dialect of Marquina*. Bilbao: Labayru Ikastegia.
Rotaetxe, K. (2008): “Lengoaiaren jatorria eta hizkuntzen sailkapenak”. In: *Jean Harischelhar-i omenaldia* (*Iker* 26). Bilbao: Euskaltzaindia. P. 577-596.
Sagarna, A. & Lakarra, J. A. & Salaberri-Zaratiegi, P. (eds.) (2011): *Euskaltzaindiaren XVI. Biltzarra. Pirinioetako Hizkuntzak. Lehena eta oraina* (*Iker* 26). Euskaltzaindia.
Sagart, L. (1999): *The Roots of Old Chinese*. Amsterdam, Philadelphia: John Benjamins.
Salaberri-Muñoa, P. (2000): “Axularren testu-zorrak”. In: *FLV*. Nr. 84. 193-210.
Salaberri-Zaratiegi, P. (1994): *Eslaba aldeko euskararen azterketa toponimiaren bidez*. Bilbao: Euskaltzaindia.
------- (1998): “Arabako mugako nafar hizkeren inguruan”. In: Camino, I. (ed.). P. 49-87.
------- (1999): “(Rewie of) Orpustan, J.B. La Langue Basque au Moyen Âge (IXe-XVe siècles)”. In: *RIEV*. Nr. 44. P. 205-209.
------- (2002): *Iraupena eta lekukotasuna. Euskal literatura idatzia 1900 arte*. Donostia-San Sebastián: Elkar.
------- (2011a): “Sobre el sufijo occidental *-ika* y otras cuestiones de toponimia vasca”. In: *FLV*. Nr. 113. P. 139-176.
------- (2011b): “Pirinioetako euskal toponimoak: direnak eta diratekeenak”. In: Sagarna, A. & Lakarra, J. A. & Salaberri-Zaratiegi, P. (eds.). P. 1019-1048.
Salaburu, P. & Lakar, M. (2005): *Baztango Mintzoa: gramatika eta hiztegia*. Pamplona-Iruñea: Government of Navarre, Euskaltzaindia.
------- & Goenaga, P. & Sarasola, I. (eds.) (2011): *Sareko Euskal Gramatika (SEG) University of the Basque Country* [available at http://www.ehu.es/seg/gizt/5, accessed on 2012-09-21].
Salazar, B. (2001): *Ugaoko euskara*. Bilbao: Ediciones Beta.
Salmons, J. C. & Joseph, B. D. (eds.) (1998): *Nostratic: Shifting the Evidence*. Amsterdam: John Benjamins.
Santazilia, E. (2010): “Un retrato del artículo vasco en el año 1895, mediante el programa VDM”. In: Aurrekoetxea, G. & Ormaetxea, J. L. (eds.). P. 175-205.

Santos-Yanguas, J. (2009): *Los tiempos antiguos en los territorios pirenaicos*. Vitoria-Gasteiz: University of the Basque Country.
Sarasola, I. (1976): *Historia Social de la Literatura Vasca*. Madrid: Akal.
------- (1980): "*Nire/neure*, *zure/zeure* literatur tradizioan". In: *Euskera*. Nr. 25. P. 431-446.
------- (1990 [1984a]): *Contribución al estudio y edición de textos arcaicos vascos* (reed. as Supplements of *ASJU* 11). Donostia-San Sebastián [= *ContTAV*].
------- (2005 [1984b]): *Euskal Hiztegia*. Donostia-San Sebastián: Elkar.
------- (1997): "Euskal hitz altxorraz". In: *ASJU*. Nr. 31. P. 617-642.
Satrústegui, J. M. (1987): *Euskal Testu Zaharrak. I*. Bilbao: Euskaltzaindia.
Saussure, F. (1916): *Cours de linguistique générale*. Paris: Payot.
Schuchardt, H. (1972 [1893]): "Sobre la formación de las flexiones de relación del verbo vasco" (translated from the original German version *Baskische Studien I: Über die Entstehung der Bezugsformen des baskischen Zeitworts*). In: *BAP*. Nr. 28(2-3). P. 217-337.
------- (1907): "Explications des formes verbales simples *dautza* et *dauntza*". In: *RIEV*. Nr. 1. P. 154.
------- (1947 [1923]): *Primitiae Linguae Vasconum*. Salamanca: CSIC.
Schuh, R. G. (1977): "West Chadic verb classes". In: Newman, P. & Newman, R. M. (eds.). P. 143-166.
------- (2003): "A comparative study of West Chadic verb suffixes". In: Bender, M. L. & Takacs, G. & Appleyard, D. L. (eds.). P. 71-86.
Siewierska, A. (2004). *Person*. Cambridge University Press.
Singh, R. (ed.) (2004). *The Yearbook of South Asian Languages and Linguistics*. Berlin: Mouton de Gruyter.
Smith, J. (2005): *Phonological augmentation in prominent positions*. London: Routledge.
Solnit, D. B. (2006): "Verb serialization in Eastern Kayah Li". In: Aikhenvald, A Y. & Dixon, R. M. W. (eds.). P. 144-159.
Soziolinguistika Klusterra (2012). *Euskal Herriko Kale Erabileraren VI Neurketa* [available at http://www.soziolinguistika.org, accessed on 2012-11-14].
Stassen, L. (2000): "AND-languages and WITH-languages". In: *Linguistic Typology*. Nr. 4. P. 1-54.
Tejerina, B. (1992): *Nacionalismo y Lengua*. Madrid: Siglo XXI.
Thieroff, R. & Ballweg, J. (eds.) (1994): *Tense Systems in European Languages*. Tübingen: Max Niemeyer Verlag.
Tovar, A. (1959): *El euskera y sus parientes*. Madrid: Ediciones Minotauro.
Trask, R. L. (1977). "Historical Syntax and Basque Verbal Morphology: Two Hypotheses". In: Douglass, W. A. & Etulain, R. W. (eds.). P. 203-217.
------- (1979). "On the Origins of Ergativity". In: Plank (ed.). P. 385-404.
------- (1981): "Basque Verbal Morphology". In: *Bascologists International Meetings* (Lejona 1980) (*Iker* 1). Pamplona-Iruñea: Euskaltzaindia. P. 285-304.
------- (1985): "On the reconstruction of Pre-Basque Phonology". In: Melena, J. L. (ed.). P. 885-891.
------- (1990): "The *-n* class of verbs in Basque". In: *Transactions of the Philological Society*. Nr. 88. P. 111-28.
------- (1995a). "Origin and Relatives of the Basque Language: Review of the evidence". In: Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.). P. 65-99.

------- (1995b): “On the History of the Non-Finite Verb Forms in Basque”. In: Hualde, J. I. & Lakarra, J. A. & Trask, R. L. (eds.). P. 207-234.
------- (1997): *The History of Basque*. London, New York: Routledge.
------- (1998): “The typological position of Basque: then and now”. In: *Language Sciences*. Nr. 20. P. 313-324.
------- (2003): “The Noun Phrase: nouns, determiners, and modifiers; pronouns and names”. In: *GB*. P. 113-170.
Túrrez, I. & Arejita, A. & Isasi, C. (eds.) (1998): *Studia Philologica in honorem Alfonso Irigoien*. Bilbao: University of Deusto.
Txillardegi, G. & Aurrekoetxea, G. (coord.) (1983): *Euskal dialektologiaren hastapenak*. Sommer University of the Basque Country.
Udolph, J. (ed.) (2012): *Europa Vasconica – Europa Semitica? Kritische Beiträge zur Frage nach dem baskischen und semitischen Substrat in Europa* [= *Beiträge zur Lexikographie und Namenforschung*. Nr. 6].
Uhlenbeck, C. C. (1909/10 [1903]): “Contribution à une phonetique comparative des dialectes basques. I, II” (translated from the original version in German: “Beiträge zu einer vergleichenden Lautlehre der baskischen Dialekte”). In: *RIEV*. Nr. 3. P. 465-503; *RIEV*. Nr. 4. P. 65-118.
------- (1947 [1942]): “Les couches anciennes du vocabulaire basque” (translated from the original Dutch by G. Lacombe). In: *Eusko-Jakintza*. Nr. 1. P. 543-581.
Ulibarri, K. (2008): “Orozkoko euskararen azterketa diakronikorantz hurbiltze bat”. In: *ASJU*. Nr. 42. P. 191-207.
------- (2010): “*Viva Jesus* dotrina: edizioa eta azterketa”. In: *ASJU*. Nr. 44. P. 41-154.
------- (2011): “1619. urtean Sevillan argitaraturiko bertsoak: *Cancion en lengua vascongada o vizcayna que contiene la sustancia deste sermon*”. In: *ASJU*. Nr. 45. P. 361-385.
Untermann, J. & Villar, F. (eds.) (1993): *Lengua y cultura en la Hispania prerromana* (*Actas del V Coloquio sobre Lenguas y Culturas Prerromanas de la Península Ibérica*, Cologne 1989). University of Salamanca.
Urgell, B. (1986): “Egiaren kantaz: I. testua eta iruzkinak”. In: *ASJU*. Nr. 20. P. 75-148.
------- (1987): “Esku-liburuaren grafi aldaketak (1802-1821)”. In: *ASJU*. Nr. 21. P. 354-387.
------- (1991): “Axular eta Larramendi”. In: *ASJU*. Nr. 25. P. 901-928.
------- (ed.) (2001): *Gueroco Guero*. Bilbao: Euskaltzaindia.
------- (2006): “Para la historia del sustantivo verbal en vasco”. In: Lakarra, J. A. & Hualde, J. I. (eds.). P. 921-948.
------- (2008): “Adizki jokatugabeak”. In: *Euskal Gramatika Historikoa*. Unpublished manuscript. University of the Basque Country.
Urquijo, J. de (1908/09): “Los refranes vascos de Sauguis traducidos y anotados”. In: *RIEV*. Nr. 2. P. 677-724; *RIEV*. Nr. 3. P. 144-157.
Urquizu, P. (2000): *Historia de la Literatura Vasca*. Madrid: UNED.
------- (2008-09): “Un texto del siglo XVII: *SUMARIO BREBEA...* (Azkoitia, 1614)”. In: *Revista de lenguas y literaturas catalana, gallega y vasca*. Nr. 14. P. 207-217.
Valcárcel, V. (ed.) (1995): *Didáctica del latín. Actualización científico pedagógica*. Madrid: Ediciones Clásicas / Colegio de Doctores y Licenciados de Vizcaya.
Velaza, J. (2012): “El vasco antiguo y las lenguas vecinas según la epigrafía”. In: Igartua, I. (ed.). P. 75-84.

Vennemann, T. (1974): “Topics, Subjects, and Word Order: From SXV to SVX via TVX”. In: Anderson, J. M. & Jones, C. (eds.). P. 339-376.
------- (1994): “Linguistic reconstruction in the context of European prehistory”. In: *Transactions of the Philological Society*. Nr. 92. P. 215-284.
------- (writings edited by Noel Aziz Hanna, P.) (2003): *Europa Vasconica – Europa Semitica* (Trends in Linguistics. Studies and Monographs 138). Berlin: Mouton de Gruyter.
Villar, F. & Prósper, B. (2005): *Vascos, celtas e indoeuropeos. Genes y lenguas*. University of Salamanca.
------- & Encarnaçao, J. D. (eds.) (1996): *La Hispania prerromana. VI Coloquio sobre lenguas y culturas prerromanas de la Península Ibérica*. Universities of Salamanca and Coimbra.
------- & Fernández-Álvarez, M. P. (eds.) (2001): *Religión, lengua y cultura prerromanas de Hispania* (*Actas del VIII Coloquio sobre Lenguas y Culturas Prerromanas de la Península Ibérica*, Salamanca 1999). University of Salamanca.
Villasante, L. (1979 [1961]): *Historia de la literatura vasca*. Oñati: Aránzazu.
------- (1994): “75º aniversario del congreso de Oñati (1918-1993)”. In: *RIEV*. Nr. 39. P. 77-96.
Yrizar, P. (1973): “Los dialectos y variedades de la lengua vasca”. In: *BAP*. Nr. 29. P. 3-78.
------- (1981): *Contribución a la dialectología de la lengua vasca* (2 vols.). Donostia-San Sebastián: Caja de Ahorros Provincial de Guipúzcoa.
------- (1991-2008): *Morfología del verbo auxiliar vasco* (15 vols.). Bilbao: Euskaltzaindia.
Zabala, I. (2003): “Nominal Predication: copulative sentences and secondary predication”. In: *GB*. P. 426-447.
Zabaltza, X. (2011): *Augustin Xaho (1811-1858) Precursor incomprendido*. Vitoria-Gasteiz: Basque Government.
Zarabozo, C. (1972): “Two Borrowed Sound Changes in Basque”. In: *ASJU*. Nr. 6. P. 174-182.
Zavala, R., 2006, “Serial verbs in Olutec (Mixean)”. In: Aikhenvald, A. Y. & Dixon, R. M. W. (eds.). P. 273-300.
Zelaieta, E. (2008): *Baztan-Bidasoako hizkeren azterketa dialektologikoa*. Pamplona-Iruñea: Government of Navarre.
Zuazo, K. (1988). *Euskararen Batasuna*. Vitoria-Gasteiz: Euskaltzaindia.
------- (1998a): “Arabako euskara”. In: Knörr, H. & Zuazo, K. (eds.). P. 127-189.
------- (1998b): “Euskalkiak, gaur”. In: *FLV*. Nr. 78. P. 191-233.
------- (2006): *Deba ibarreko euskara. Dialektologia eta tokiko batua*. Oñati: Badihardugu Euskara Elkartea.
------- (2008): *Euskalkiak. Euskararen dialektoak*. Donostia-San Sebastián: Elkar.
------- (2010a): *El euskera y sus dialectos*. Zarautz: Alberdania.
------- (2010b): *Sakanako euskara. Burundako hizkera*. Pamplona-Iruñea: Government of Navarre.

General Index

A

Abaitua, J., 81, 84-86, 337
Adour River, 57
Aescoa, Valley of, xxii, xxiii, 52, 55, 56, 58, 67, 69, 296, 304
  Variety of, 55
Aescoan (subdialect of West.-LN), xxii, 8, 55, 67, 68, 80, 108, 135, 285
aetzak, 56
African languages, 189, 417
Agud, M., 186, 188, 266
Aguirre, J. B., xvii, 106
Aikhenvald, A. Y., 4, 193, 195, 417, 424, 426
Ainhoa, xviii
  Variety of, 52
Aizkibel, J. F., 108, 177
Alavese Plain, 46, 85
Albanian dialects, 83
Albelda, Monastery of, 96
Alberdi, J., 301, 360, 400, 406
Aldai, G., 47, 63, 323, 349, 360, 377, 389, 412, 424, 426, 443
Aldaieta, Archeological Site of, 85
Alfaro, E., 94
Altube, S., 444
Altuna, P., xxv, 288, 289
Amazonian languages, 220
Amerindian languages, 189, 194, 206, 218
Amescoa, Valley of, xxi, 46, 76, 453
Amilleta, J. B., xvii
Amonarriz, K., 24
Amorebieta-Zornotza, 77
Amorrortu, E., 6, 11, 18, 20, 25, 26
Anderson, G. D. S., 416
Andia, Mountain Range of, 47, 49
Andoain, xix, 77
Angelu, 53
Añibarro, P. A. de, xvii, 106, 108, 109
Ansó, Valley of, 116
Antzuola, xvii
Anue, Valley of, 47, 294
Apalauza, A., 293
Apraiz, O. de, 116
Aquitania, 81, 84, 92
  Aquit. filiation, 81-86, 93, 146, 165, 418
  Aquit. inscriptions, 1, 82, 97, 146, 228, 275, 357
  Aquit. language, vii, xxiii, 1, 81, 87, 92, 93, 176, 185, 186, 203, 275
  Aquit. period, 175
  Aquit. population, 81, 83
Arabian language, 180
Aragón, 71, 116
  Aragonese dialect, 168, 267
Araitz (subdialect of North.-HN), 47
  Valley of, 47, 49, 69
Arakil (subdialect of North.-HN), 47
  Valley of, 47, 49, 67, 68, 294
Aramaio, xvii, xxii, 45, 46, 288, 290-292, 341
Arana y Goiri, S. de, 110
Arana, A., 59, 66, 290
Arana, J. I., 110, 111
Aranbillaga, J. d', 103
Arantza, 33, 34, 48
Arantzazu, Congress of, xxiii, 112
Araquistain, J. de J. M., 108
Arbelaiz, J. J., 178, 180, 186
Arberoa, Valley of
  Variety of, 57
Archaic Basque, 98, 99, 101, 205, 357
Archu, J.-B., 107, 108
Ardanaz, P. de, 104
Arejita, A., 314
Aresti, G., 112
Aretxabaleta, 39
Argaignarats, P., xvii
Ariztimuño, B., 7, 41, 44, 54, 143, 173, 195, 213, 283, 299, 311, 359
  origin of the ptv., 277-280
Arkotxa, A., 100
Arrangoitze
  Variety of, 52
Arrasate-Mondragón, xix, 256, 273
Arratia, Valley of
  Variety of, 35
Arriandiaga, M., 110
Arrigorriaga
  Variety of, 35
Arriolabengoa, J., 100
Artiagoitia, X., 149, 163, 175, 233, 234, 252, 264
Artzibar, Valley of, 56

Variety of, 50
Arzadun, M. de, 103
Arzamendi, J., 97
Ascoli, Bronze of, 82, 84
Asón River, 70
Astarloa, P. P., 109, 360
Asteasu, xvii, 106
Astigarraga, L., 108
Atetz, Valley of, 47, 294
Atharratze, 77, 78
Variety of, 59
Atxaga, B., 438
Auritz-Burguete
Variety of, 50
Aurrekoetxea, G., xxv, 304, 367, 399, 406, 407
Australian languages / language family, 189, 194, 196, 198, 206, 261
Austro-Asiatic language family, 190
Austronesian languages / language family, 173, 180, 182, 205
Autrigones, 70
Awutu language, 196
Axular, P. de, xvii, 103, 104, 288, 309, 310, 315, 415, 421
Ayala, Valley of, 35, 114
Aymaran languages, 28
Azkarate, A., 85
Azkarate, M., xxv, 288, 289
Azkoitia, xxi, 8, 102, 266, 294
Azkue, R. M. de, xvii, xxii, xxv, 45, 55, 60, 68, 69, 110, 111, 113, 141, 177, 186, 228, 267, 271, 291, 301, 302, 308, 329, 391, 392, 444
Azpeitia, xix, 44, 102, 294, 320
Variety of, 43

B

Bähr, G., 111
Baigorri (subdialect of West.-LN), 55
Baigorri, Valley of
Variety of, 55
Bakio, 302
Bakrò-Nagy, M. Sz., 180, 182, 189
Baldi, P., 115
Bantu languages / language family, 194, 195, 330
Baranbio, 302
Barbier, J., 110
Bárdenas Reales, Region of, 61
Bardoze, 57, 129
Barkoxe, 78
Baroja, S., 110
Barreña, A., 6, 11
Barrena, E., 72
Barrutia, P. I. de, xvii, 104, 311
Bartoli's principle, 113
Basabürüa, Area of, 60
Basaburua, Valley of, 47, 49, 69
Basque Autonomous Community, 10, 11, 31, 35, 74, 283
Basque Statute of Autonomy, 17, 112
Basque-Caucasian comparison, 177, 183
Basque-Iberian question, 1, 5
Baxter, A. N., 196
Bayonne, 53
Diocese of, 72
Baztan (subdialect of North.-HN), 47, 51
Valley of, xvii, xxii, 44, 47, 49, 52, 54, 56, 68, 69, 165, 287, 293, 304
Bearn, 60, 94, 328
Bearnese language, 132, 168, 328
Bedaxagar, M., 223
Bela, J. de, 100, 105
Belapeyre, A. de, xviii, 102, 290
Belate, Pass of, 47
Benveniste, É., 178, 179, 180, 207, 212
Bera, 33
Variety of, 47
Berber languages / language family, 175, 229
Bergara, xxi
Variety of, 35
Beriayn, J. de, xvii, 51, 102, 103
Bermeo, 41, 46, 77, 131, 187
Variety of, 35
Bernat, F., 60
Berones, 70, 71
Bertso Bizkaitarrak, 104
bertsolaris, 24, 109
Beskoitze, xx, 8, 67
Variety of, 57
Beterri, Area of, 43, 47, 76, 77
Betolatza, xvii
Betolaza, J. Perez de, xvii, 101
Biarritz, 53, 134
Miarritze, 134, 340
Bidankoze, 60
Variety of, xxii, 59, 60

Bidarte, xxi
Bidasoa River, 47, 293
Bidégaray, D., 105
Bigorre, 94
Bilbao, xvii, xx, xxi, 8, 35, 43, 77, 105, 114
Bilintx (pseudonym of G. J. I. Bizkarrondo), 109, 110
Biscayan from Guipuscoa (subdialect of B), 35
Biscayan-Basque Theory, 73
Blevins, J., 119, 127, 130, 161
Blust, R. A., 180, 182
Boas, F., 184
Böhme, H. W., 85
Bonaparte, L.-L., xviii, xxii, xxiii, 34, 35, 70, 80, 86, 108, 113, 114, 187, 202, 285, 304, 400
  Al, 45
  B, 35
  East.-LN, 57, 58, 59
  G, 43
  L, 52
  North.-HN, 46, 47
  phonetics, 134, 151
  S, 59, 60
  South.-HN, 49-51
  trips and classifications, 66-69
  verb morphology, 360
  West.-LN, 55, 56
Bongo-bongo linguistic method, 183
Bortziriak-Cinco Villas (subdialect of North.-HN), 47, 49, 51
  Area of, 33, 47, 48
Botorrita (Zaragoza), 82
Broussain, P., 73
Bueno, A., 447
Burgi, 60
Burunda, Valley of, xxi, xxii, 39, 43, 46, 49, 67, 72, 76, 79, 249
  Variety of, 43
Bustos, J. J. de, 97
Bybee, J. L., 4, 197
Byrne, F., 196

C

Caesar, Julius, 92
Calagurris, 115
Calahorra, Diocese of, 72, 76
Camino, I., 31, 34, 49-52, 55, 56, 58, 83, 285, 286, 289, 293-296, 304, 305
  dialectal branching, 78
Caminreal (Teruel), 5
Campbell, L. R., 93, 445, 449
Campión, A., 50, 68, 70, 71, 73, 108
Capanaga, M. Ochoa de, xviii, 103, 108, 292
Cardaberaz, A. de, xviii, 106, 107, 444
Caristii, 70, 71, 81, 83, 84, 113
Caro-Baroja, J., 33, 71
Carolingian Renaissance, 96
Castaños-Garay, F., 227, 229, 232, 235-239, 244, 254
Castile, 96, 264
Castilian
  C. cultural sphere, 104
  C. language, 21, 35, 45, 100
Caucasian languages, 178, 189, 193
Celtiberian language, 94
Celtic languages / language family / filiation, 71, 82-84, 86, 93, 115, 185
Cerrón-Palomino, R., 32, 33
Chaho, J.-A., 107, 109, 110
Charles III (King of Spain), xix
Chartres, B. of, 3
Chinese language, 180, 196, 203, 395, 422, 459
Cidacos River, 92
Cizain-Mixain (subdialect of East.-LN), 57
Cize, Valley of, xviii, 33, 58, 60, 67, 72, 302
  Variety of, 57
Classical Souletin, xx
Comrie, B., 423, 445
consecutio temporum, 64
Contestani, 5
Coptic, 94
Corbett, G. G., 4
Corominas, J., 116, 186, 213, 216
Counter-Reformation, 101, 102
Creissels, D., 227, 255
Cushitic languages, 195

D

d'Albret, J. (Queen of Navarre), 101
Danish language, 122, 176
Dassança, M., 103
Dax, Diocese of, 72

Deba
  Deba River, 70
  High Deba, 39, 46
  Town, 46
  Valley of, xxii, 8, 35, 70, 80, 294, 338
Dechepare, B., xviii, xxii, 90, 98, 100, 101, 106, 160, 229, 268, 269, 288, 311, 343, 349, 350, 390, 402, 409, 415
Diharce d'Alçuyet, G., xix
Dixon, R. M. W., 4, 189, 240, 242, 423, 424, 444
Doctrina Christiana (of Zegama), xviii
Dodgson, E. S., 107, 108, 117
Domene, J. F., 233
Donapaleu, xx, 77
Donegan, P., 130, 178, 191, 198, 201, 204, 457
Doneztebe, 47
Donibane Garazi, xviii, 57, 60, 77, 245
Donibane Lohizune, xix, xxi, 103, 339
  Variety of, 52
Dryer, M. S., 458
Duanmu, S., 203
Duhalde, M., 107
Durango, 46, 77, 103
Durie, M., 178
Duvoisin, J. P., xviii, 73, 107, 108

E

Early Modern Basque, 105
Eastern Biscayan (subdialect of B), 35
Eastern Low Navarrese of the Adour (subdialect of East.-LN), 57
Eastern Southern High Navarrese (subdialect of South.-HN), 50
Ebro River, 83
Echave, B. de, 100
Echegaray, C. de, 72
Echenique, M. T., 95
Egiaren Kanta, 104
Egues, Valley of, 51
  Variety of, 50
Eguiategui, J., 107
Eguren, J. M. de, 108
Egurtzegi, A., 6, 119, 122, 126, 127, 137, 140, 143, 146, 159, 160, 173, 203
Eguzkitza, A., 314, 315
Eguzkitza, J. B., 111
Eibar, xx, 41, 302
Eliçalde, F. de, xviii, 51, 102
Elissamburu, J.-B., 109
Elizondo
  Variety of, 47
Elkano, xx
Elordieta, A., 168, 288, 312
Elordieta, G., 119, 158, 168, 170, 171, 178, 288, 296, 312
Elso, S. de, 101
Emilian Glosses, 90, 97, 98, 302, 353
enarak, 61
Encartaciones, Area of, 35
English language, 189, 217, 302, 315, 362, 397, 416, 453
  Old English, 416
Ensunza, A., 448
Epps, P., 218, 220
erdaldunas, 19
erdara (or non-Basque language), 143, 444
Erdozia, J. L., 49, 304
ergative displacement, *see* split ergativity
Ergoiena, Valley of, 43, 49
Ernout, A., 213
Erro, Valley of
  Variety of, 50
Esnaola, I., 24
Esquiule, 60, 343
Estella-Lizarra, 82
Etchahun, P. T., 109
Etchepare, J., 110
Etcheverry de Ciboure, J., xviii, 103, 105
Etcheverry de Sara, J., xix, 105, 108
Etcheverry, P. d', 103
Etxalar, 34
Etxarri-Aranatz, 43, 47, 49, 68, 142
  Variety of, 43
Etxeberria, U., 278
Etxepare, R., 240
Euri-dacar, 115
Euskal Herria, 11
euskaldun berris, 19, 22, 35, 103
euskaldun zaharras, 19, 22, 26
Euskaltzaindia, xvii, 111, 112, 116
euskara batua, 26, 27, 32, 34, 112, 284, 324, 326, 361, 447
Eusko Ikaskuntza, 111
Ewe language, 196
Eys, W. J. van, 108, 244, 270, 400
Ezkabarte, Valley of, 294

F

Fant, G., 130
Feng, S., 203
Ferrer i Jané, J., 5
Finnish language, 277
Fita, F., 107, 108
Foral Community of Navarre, 11
Frajzyngier, Z., 193, 195
Franco era/regime, 16, 22
Franco-Aquitanian filiation, 85
Fray Bartolomé, xix, 106
Frellesvig, B., 201
French Basque Country, 11, 12, 14, 18
French Enlightenment, xxi, 107
French language
  influence of the, 24, 105, 127, 328, 383
  linguistic phenomena of the, 62, 128, 302, 355, 402, 416, 444
French Revolution, 328
Fueros, Loss of the Basque, 109

G

galdegaia, 310, 440, 441
Gallo-Romance languages, 128, 129
Gaminde, I., 64, 288
Gamiz, J. B., xix, 45, 106
Gamkrelidze, T. V., 182
García de Albeniz, J., 42, 106
Garcia de Salazar, L., 99
Garde, 60
Gares
  Variety of, 50
Garibay, E. de, xix, 98-100, 127, 273, 274
Garonne River, 83, 86, 94
  High Basin of, 12, 81-83, 93
Garrett, A., 161, 194
Gascon language / linguistic domain, xxiv, 53, 127, 129, 130, 153, 187, 200-202, 328, 355, 382
Gasteluçar, B., xix, 103
Gatika, 64, 288, 290, 291
Gaulish language, 94
Gautegiz-Arteaga, xx
Gavel, H., 126, 128, 129, 141, 166, 232, 238, 244, 257, 263, 381, 394
General Privilege of Navarre, 98
German language, 355, 397
Germanic languages / linguistic tradition, 2, 115, 265
Gernika, 77, 125
  Variety of, 35
Gers, Plain of, 94
Gesalatz, xviii, 49, 51
Getaria, 85
  Guetaria (of Guipuscoa), 85
  Guétary (of Labourd), 86
Getxo, 39, 136, 137
Gèze, L., 108, 257
Gilligan, G., 197, 444, 447
Goenaga, P., 378
Goierri, Valley of, 39, 142
Goizueta, 166, 168, 294
Gómez, R., 60, 177, 189, 192, 195, 196, 239, 267, 270, 326, 360, 430, 454, 456, 458
Gómez-Moreno, M., 5
Gómez-Pantoja, J., 94
Goñerri, Valley of, 51, 286, 294
González-Perujo, J. M., 94
Good, J., 194
Gorrochategui, J., 4, 10, 31, 71, 81-85, 92-95, 102, 185, 186, 197, 283
Goyetche, L., 107
Goyheneche, E., 116
grammaticalization processes
  3rd grade demonstrative (**har*) > Article (*-a*), 228, 234, 294
  Agentive complement (**-ga*) > ERG (*-k*), 242, 459
  ALL + Come + Move (**ra-din-non*) > TERM ALL (*-raino*), 269
  ALL + Face (*-ra-antz*) > DIRECTIONAL ALL (*-rantz*), 267
  ALL + L.GEN (*-ra-ko*) > DST (*-rako*), 273
  ALL + Put, sit (*-ra* + **dar*) > DIRECTIONAL ALL (*-ra-t*), 261
  ALL + Receptacle (**-ra-ontz*) > DIRECTIONAL ALL (*-rantz* / *-rontz* / *-runtz*), 267
  ALL + Top + Up to (**-ra-gain-do*) > TERM ALL (*-raino*), 268
  Animacy marker in oblique NPs (*-ga-*) > ERG (*-k*), 241
  Article (*-a*) > Singulative marker (*-a*), 233
  Be (*egon*) > PROG (*irakurtzen dago*), 350, 383
  Be engaged on (*jardun*) > PROG (*egiten dihardu*), 350

Be held (**har-i*) > PROG (*egiten ari naiz*), 349, 412
Become (**din*) > derivational suffix (*-ti*) > ABL/PROS (*-ti(-ka)*), 265
Carry (*eroan/eraman*) > HAB (tr.) (*egin daroa/darama*), 350, 413
Cling to (**duts*) > ditrans. aux. (*-u(t)s-*), 41
Come (**din*) + adv. *-ka* > ABL-PTV (*-(r)ik*), 277
DIRECTIONAL ALL + Move + *-tz* (**-rada-non-tz*) > DIRECTIONAL ALL (*-rantz* / *-rontz* / *-runtz*), 268
Do (*egin*) > verbal FOCUS (*jauzi egin da*), 348, 443
Finish (**den*) > 3rd p. pl. A (*-de*), 196, 238, 371
Finish (**den*) > dat. pluralizer (*-e-i*), 245
Finish (**den*) > PFV (*-en*), 196, 238, 417, 424, 426
GEN (*-(ar)en*) > DST (*-(ar)en*), 273
GEN (*-ko/-(r)en*) > PROSP (*egin-go / egin-en dut*), 345
GEN + Company + INES (**-(ar)e-kide-n)* > SOC (*-(ar)ekin*), 269
GEN + L.GEN INDEF (**-(ar)en-tako*) > DST (*-(ar-)endako*), 272
GEN INDEF (*-ren*) > ABL (*-rean*), 266
Give (**nin*) + adv. *-ka* > ABL-PTV (*-(r)ik*), 277
Give (**nin*) > aux. of ethical dat. (*-i-*), 196, 404
Give (**nin*) > DAT (*-i*), 245, 390, 427
Give (**nin*) > ditrans. aux. (*-i-*), 196, 246, 390, 426
Go (*joan*) > HAB (intr.) (*galdu doa*), 383, 413
Know (*jakin*) > HAB (*badakie etorten*), 413
L.GEN (*-ko*) > DST (*-ko*), 273
L.GEN INDEF (*-tako*) > ESSIVE/TRANSL (*-tako*), 276
Locative (*-(t)za*) > INSTR (*-z*), 247
Locative (**dar*) > 1st p. sg. A (*-da*), 366
Locative (*-ga*) > 2nd p. sg. A (masc.) (*-k*), 367
Locative (*-ga*) > plural marker (*-k*), 232, 367
Locative + Put, sit (**-(t)za-da*) > ESSIVE/TRANSL (*-tzat*), 276
Now (**da(n)*) > PRS (*da-*), 259, 375
Palm (of the hand) (*ahur*) > 1st grade demonstrative (*hau(r)*), 299
Put, sit (**dar*) > 1st p. sg. A (*-da*), 366
Put, sit (**dar*) > ALL (*-ra*), 260
Put, sit (**dar*) > IPFV (*da-*), 193, 196, 260, 375, 417, 424
Same (*ber*) > 3rd p. pron. (*ber-*), 317
Take (**don*) > CAUS (*-ra-*), 193, 417, 424, 427
Take (*har*) > 3rd grade demonstrative (*har-*), 195, 229, 299
Thing + PROS (*-gai-ti*) > MOT (*-ga(i)ti(k)*), 272
Top (*-gan*) > INES (*-an*), 250
Vbal. noun + INES + Be (*joa-te-n da*, **egi-te-n da*) > PROG > GNOMIC ~ HAB (*joaten da* / *egiten du*), 326, 416
Walk (*ibili*) > PROG (*egiten dabil*), 350, 383
Grassmannian dissimilation, 78, 152, 160, 214
Greek language, 115, 152, 397
Greenberg, J., 189, 191, 239, 453
Grosu, A., 438
Guipuscoan from Navarre (subdialect of G), 43
Guipuscoan Navarrese (subdialect of North.-HN), 47
Guiter, E., 96
Gulina, Township of
Variety of, 50
Gussenhoven, C., 168

H

Haase, M., 355
Hamp, E., 183
hapax, 94
Haranburu, J., 103
Haraneder, J., xix
Harff, Arnold von, 98, 100, 230
Harizmendi, C., 103
Harriet, M., 109
Harriet, M. de, 105, 360
Harris, A. C., 183, 445
Haspelmath, M., 363
Hautes-Pyrénées, Département des, 82
Hawkins, J. A., 197, 438, 443, 444, 447
Hazparne, xix, 67
Heath, J., 371, 377
Hebrew language, 180
Heine, B., 4, 193, 246, 247, 260, 277, 375, 424, 427
Hengeveld, K., 458
Henry IV (King of France), 102
Hernani, xviii
Variety of, 43, 77

Hidalgo, B., 444
Hiriart-Urruty, J.-B., 110
Hiribarren, J. M., 107
Hiriburu, 67
Hittite language, 456
Hokan languages / language family, 395
Holmer, N. M., 189, 201
Hondarribia, 33, 72, 187
Houis, M., 189
Hoz, J. de, 5
Hualde, J. I., v, vii, xxv, 4, 119, 170, 288, 296, 312
  /-wh-/ > /-f-/, 153, 176, 216
  figure of Bascology, 3
  old accentuation, 170, 171
  old stops, 122, 176
  phonetics and phonology, 120, 126, 128, 129, 131, 133-136, 141, 153, 154, 158, 163, 164, 169
  study of accents, 59, 166-168
Huesca, 97
Humboldt, W. v., 107, 192, 236, 454
Hungarian, 1
Hyman, L., 195, 197, 204
hyper-Biscayan, 110

I

Ibargüen-Cachopín, Chronicle of, 98, 99
Ibarra, O., 24, 47-49, 52, 293, 304
Iberian language, 5, 70, 82, 196, 219
Igartua, I., 126, 144, 152, 170, 175, 223, 247
Igartua, M., 257
Iggesen, O. A., 225
ikastolas, 16, 17, 18
Imotz, Valley of, 47, 49, 293
Indo-European
  IE. filiation, 82, 84, 95, 115, 179, 180
  IE. languages / language family, 93, 95, 115, 173, 178, 179, 182, 183, 190, 212, 221, 247, 359, 381
  IE. linguistic tradition, 2, 4, 115, 160, 180, 205
  Indo-Europeization, 74
Indo-Hittite theory, 73
Indo-Iranian languages, 152, 456
Indosphere, Languages of the, 191
Intchauspe, M., 108, 400
internal reconstruction method, vii, xxiii, 8, 90, 95, 174, 177, 297, 299, 321, 417, 429, 455, 460
Intza
  Variety of, 47
Inza, D., 43
Iparraguirre, J. M. de, 109
Irantzu, Monastery of, 96
Iraola, V., 110
Iratxe, Monastery of, 96, 97, 243, 303, 450, 451
Irigoyen, A., 97, 214, 229, 243, 287, 290, 294, 296-299, 320, 321, 413
  demonstrative > article development, 65, 229, 241
  rise of the 3-grade deictic system, 296-298
Irun, 72, 259
  Variety of, 47
Iruña-Veleia
  affair of, 93, 94
  archeological site of, 83
Italian language, 416
Ito, S., 246
Iturriaga, A., 107
Itziar, 44
Ivanov, V. V., 182
Iverson, G., 175
Izaba, 60, 168, 169
Izarbeibar (subdialect of South.-HN), 50
  Valley of, 51
Iztueta, J. I. de, 106, 107, 109

J

Jacobsen, W. H., 249, 250, 252, 254, 268, 367, 368, 371
Janhunen, J., 200
Japanese language, 437
  Proto-Japanese, 201
Jauregi, O., 185
Jáuregui, F., 108

K

Kadet eta Bettiriño, xix, 106, 107, 343
Kartvelian languages / language family, 175, 183
Kempis, T. de, xx
Kerejeta, M. J., 54
Kerongo language, 277
Kintana, X., 402

Kirikiño (pseudonym of E. Bustinza), 110
Kitson, P. R., 115
Knörr, H., 95, 100, 116
Kontrasta, 115, 453
Kristang creole language, 196
Kuryłowicz, J., 406
Kuteva, T., 4, 193, 246, 260, 277, 375, 424, 427
Kwa languages, 204

L

Lacombe, G., 73, 381, 394
Ladefoged, P., 130
Lafitte, P., 257, 335
Lafon, R., 63, 93, 260, 285, 323, 348, 360, 365, 366, 375, 377, 381, 388, 390, 426
  abl./pros., 263-265
  alloc. speech, 57, 399, 402, 409
  Caucasian comparison, 177, 193
  dat. flag *-ki*, 329, 386
  Dch.'s language, 57, 58
  figure of Bascology, 2, 3
  noun declension, 223, 224, 228, 229, 234, 236, 237, 243, 248, 254, 256, 270, 275, 277, 279
  root structure, 177, 178, 192, 206
  S. /y/, 128, 129
  verbal diathesis, 365, 382, 383, 385, 394, 395, 422
Laka, I., 110, 377
Lakar, M., 44, 49, 293, 304
Lakarra, J. A., 4, 6, 10, 31-33, 89, 92, 95, 102, 115, 119, 173, 187, 189, 201, 207, 223, 229, 238, 241, 247, 271, 281, 283, 323, 357, 359, 360, 367, 429, 446
  against Basque-Iberian words/numerals, 5, 196, 220
  aspiration (phenomena), 170, 176, 200
  Common Basque, 78, 80, 188, 202
  etymologies, 139, 140, 144, 163, 165, 177, 185, 187, 214-217, 219, 221, 235, 243, 265, 303, 389, 393, 425, 427, 452
  figure of Bascology, 3, 4
  grammaticalization of serial verbs / SVCs, 193, 195, 196, 229, 238-240, 245, 246, 260, 261, 265, 268, 269, 276, 277, 299, 328, 366, 371, 375, 389-391, 404, 417, 418, 422, 424, 426, 427
  holistic approach, 178, 190, 197, 204, 239, 326, 417, 430, 452-454, 457, 458, 460
  Old B, 42, 43, 73, 184, 261, 262, 266
  old consonant system, 62, 79, 175, 176, 197, 198, 265, 327
  philological issues, 90, 98, 101, 104, 186, 271, 335, 360
  theory of the root, 9, 124, 125, 149, 162, 163, 171, 173, 181-187, 192, 197, 201, 202, 204, 205, 233, 234, 239, 241, 252, 323, 327, 330, 391, 458
  verb morphology, 41, 139, 195, 260, 329, 331, 333, 340, 351, 354, 355, 391, 393, 422
Lana, Valley of, xxi, 46, 76
Landucci, N., xx, 100, 113, 332, 341
Lantz, Valley of, 47
Laphitz, F., 107
Lardizabal, F. I., 108, 109, 409
Larraine, 60, 168
Larramendi, M. de, xix, 103-109, 360
Larrasquet, J., 128
Larrea, xix
Late Modern Basque, 110
Late Vasconization Theory, 84, 85
Latin
  L. influence, 2, 9, 85, 91, 93, 95, 96, 110, 124, 134, 135, 141, 143, 145, 148-150, 152, 153, 155, 159, 162, 165, 175, 176, 177, 181, 236, 238, 268, 272, 289, 329, 332, 334, 345, 354, 355, 375
  L. language, xix, xxiv, 90, 94, 98, 104, 105, 115, 176
  L. linguistic phenomena, 124, 224, 225, 316, 397
Latin-Romance influence, 195, 331, 355, 356, 357
Lauaxeta (pseudonym of E. Urkiaga), 110
Laukariz, xx
Lazarraga, J. Perez de, xix, 4, 8, 100, 101, 113, 262, 310, 332, 381, 391, 402
Lécluse, F., 108
Legutio, xxii
Lehmann, W. P., 445, 457
Leiçarraga, J., xx, 8, 33, 100-104, 135, 141, 288, 290, 301, 343, 391, 402, 405, 415
Leintz-Gatzaga, Valley of, 341
  Variety of, 35, 46
Leire, Monastery of, 96
Lekeitio, xvii, 55, 288, 290-292, 295, 296, 312
León, A., 232
Lerga, Inscription of, 82, 146, 165

Lezamis, J. de, 266
Li, Ch. N., 459
Líbano-Zumalacárregui, A., 97
Liçarraga de Elcano, J., xx, 51, 106
Lichtenberk, F., 196
Ligi, xx
Linares River, 92
Linschmann-Aresti Law, xii, 7, 55, 306, 312, 315, 316, 317, 319
littérature du colportage, 109
Lizaso
  Variety of, 47
Loiola, xix
López, 106, 107
Lord, C., 4, 195, 196, 299
Lubieta, J. D. de, 105
Luchaire, D. J. A., 93
Lujanbio, O., 166, 168
Lusitanians, 84
Luzaide-Valcarlos, xxiii, 67, 69

M

Madariaga, P. de, 127
Maddieson, I., 130
Maister. M., xx
Mallory, J. P., 83
Maluquer, J., 75
Mañaria, xviii
Mañaricúa, A. de, 72
manexak, 60
Manning, C. D., 438
Manterola, José, 109
Manterola, Julen, 65, 215, 227-230, 236-239, 248, 253, 254, 256, 281, 288, 447
  article in the Middle Ages, 97, 143, 195, 228, 339
Marineus Siculus, Lucius, 98, 100, 230
Markina, xix, xx, 136, 137, 288, 290-292
  Variety of, 35
Martinet, A., 2, 96, 123, 175
  theory of stops, 122-124, 162, 174-176
  theory of the accent, 169-171, 201
Martínez de Isasti, L., 33, 99, 100
Martínez de Ochoa, S., 104
Martínez-Areta, M., vi, viii, 1, 6, 7, 31, 89, 170, 171, 173, 178, 198, 223, 283, 300, 323, 327, 359, 373, 429
  noun declension, 231, 233-236, 242, 247, 252, 266, 281
  verbal diathesis, 360, 365, 366, 377, 383, 394-397, 420, 422, 423, 425, 426
  word order, 430, 438, 458, 459
Martínez-Sáenz de Jubera, M., 94
Materre, E., 103
Matisoff, J., 144, 180
Maule, xx, 60, 61, 77, 78, 257, 263
Maytie, A. F. de, 102
Medieval Basque, 96, 97, 115, 357
Meillet, A., 89, 113, 117, 183, 189, 213
Mendiburu, S. de, xix, xx, 106
Mendieta, F. de, 99
Mendigacha, M., 60
Mendigorría (Navarre), 5
Mendionde-Lekorne, Township of
  Variety of, 55
Menéndez-Pidal, R., 72
Merino-Urrutia, J. J., 116
Michel, F., 109
Michelena, L., xxiii, xxv, 6, 81, 89, 113, 116, 124, 137, 138, 173-175, 178, 183, 184, 186, 192, 201, 453
  Common Basque, 9, 31, 73-75, 78, 86, 113, 122, 173, 188, 202, 417
  dialectal issues / classification, 32, 34, 41, 48, 49, 60, 69, 70, 73, 157, 168, 184, 202, 229
  etymologies, 138, 178, 186, 187, 195, 196, 212, 213, 215, 217, 265, 297, 303, 452
  figure of Bascology, vii, 2, 3, 5, 119, 171, 173, 174, 180
  noun declension, 65, 97, 224, 227-229, 232, 233, 236, 239, 240, 243, 247, 251, 253, 254, 257, 262-266, 270-272, 275, 276
  old accent system, 48, 151, 166, 167, 169-171, 178, 201
  old consonant system, 9, 96, 122, 123, 124, 143, 145, 148, 150, 159, 162, 163, 172, 174-178, 181, 187, 197, 199, 200, 204, 205, 207, 260
  philological issues, 46, 89, 90, 93, 97, 112, 116, 117, 149, 155, 266, 453
  phonetic processes (especially in loans), 61, 79, 95, 96, 126-128, 130-133, 135, 142, 148, 151, 158, 164, 165, 195, 200, 258
  verb morphology, 65, 195, 326, 334, 340, 345, 354, 355, 360, 388, 389, 422
Micoleta, R. de, xx, 105, 108, 155, 360
Middle Ages, 35, 45, 46, 76, 95, 96, 122, 126, 127, 132, 142, 143, 150, 186, 339, 353, 356, 427
  Early Middle Ages, 84

mintzagaia, 312
Mirande, J., 112
Miserere, 102
Mixe, Area of, xxii, 62, 67, 69, 129
  Variety of, 57
Mixean languages, 366
Mixed Labourdin (subdialect of L), 52
Mixe-Zoquean languages / language family, 371
Moguel, J. A., xvii, xx, 106, 107, 109, 212, 263
Moguel, V., 106, 107
Mon-Khmer languages / language family, 189-191, 201, 457
Montori, 78
Mounole, C., 42, 194, 195, 227, 255, 323, 342-344, 349, 351-355, 360, 366, 411, 416
Mugarza, P., 40
Múgica, S., 72
Mujika, J. A., 91, 378
Müller, F., 394
Munda languages / language family, 190, 191, 198, 201, 261, 457
  Proto-Munda, 201

N

Narrog, H., 246, 247
Nationalism, Basque, 110
Navarre, Kingdom of, xviii, 99
Navarrese-Labourdin, 53, 68, 69, 107, 291
Navarro-Tomás, T., 125, 126, 132
Neo-Grammarians, 2
neologisms, 106, 109-111, 329
Nervión River, 35, 70, 81, 86
New Caledonia(n languages), 196
New Guinean languages, 206, 417
Newman, J., 196, 260
Non-Indo-European language / filiation, 1, 16
Northern Guipuscoan (subdialect of G), 43
Novia de Salcedo, P., 109
Núñez-Astrain, L., 12, 229

O

Occitan dialects / area, xxiv, 127, 129, 328
Oceanic languages, 417
Ochoa de Arin, J., xx, 102
Odieta, Valley of, 294
Odriozola, J. M., 109
Ohala, J. J., 127, 149, 160, 161, 172
Oiartzun, xx, 72, 84
Oiasso, 115
Oihenart, A. d', xviii, xx, 100, 102, 103, 128, 149, 264, 287, 288, 297, 343, 360, 381, 409
Olabide, R., 110
Olaechea, B., xx
Olaibar, Valley of
  Variety of, 50
Old Basque, 82, 93, 94, 96, 113, 171, 175, 182, 193, 276, 357
Old Biscayan, 243, 244, 249, 255, 261, 262, 271, 325, 333, 378, 390, 414
Old Classical Basque, 102
Old Common Western Basque, 351
Old European, 115, 201
Old Guipuscoan, xx, 271
Old Irish, 456
Old Labourdin, 227, 313
Old Souletin, 311, 343
Old Spanish, 416
Olloibar, Valley of, 51, 294
Oloron, Diocese of, 72
Oltza, Township of, 51, 294
  Variety of, 50, 51
Olutec language, 196, 366, 367
Oñati, 39, 46
Ondarroa, 301
Oñederra, L., 128, 129
Ordizia, xx, 44, 102
Orduña, E., 5
Oregi, J., 375, 420
Orixe (pseudonym of N. Ormaetxea), 111, 212, 267
Ormaetxea, T., 288
Orozko, 64, 134
  Variety of, 35
Orpustan, J.-B., 97, 116
Orreaga-Roncesvalles, 276
Ortega, A., 6, 11, 18, 20, 22
Ortiz de Urbina, J., vii, xxv, 377, 423, 442
Osa-Unamuno, E., 429
Otaño, P. M., 109
Otxandio
  Variety of, 35
Oyanguren, M. de S. I., 105

Oyharçabal, B., 100, 105, 323, 328, 345-347, 378
Oztibarre, Valley of, 72

P

Padilla-Moyano, M., 7, 245, 271, 283, 291, 314, 323, 325
Page, B. R., 115
Pagliuca, W., 4, 197
Pagola, R. M., 33, 50, 55, 66-68, 70, 72, 73, 114, 285
Palancar, E. L., 247
Palay, S., 186, 219
Pallars, Valley of, 75
Pama-Nyungan languages / language family, 247
Pamplona-Iruñea, xix, xxi, xxiii, 46, 48-50, 60, 72, 75-77, 85, 286, 294, 317
  Diocese of Pamplona, 72, 76
  Kingdom of Pamplona, 75, 96
Papuan languages, 196
pardixak, 56
Pascual, J. A., 186, 213, 216
Pawley, A., 194
Peñaflorida, Count of, xxi
Penny, R,, 153
perceptual dialectology, 34
Pérez-Bilbao, A., 131
Perkins, R., 4, 197
Peterson, D., 97
philology of precision, 117
Picaud, Aimery, 98, 127, 230
Pikabea, J., 53, 54, 56, 58, 245, 291
Plentzia, 39
  Variety of, 35
Pop, S., 113
Portal, M., xxi, 46
Post, M. W., 191, 192, 203
Pouvreau, S., xxi, 103, 105
Poza, A. de, 100
Proper Labourdin (subdialect of L), 52
Proper Souletin (subdialect of S), 59
Prósper, B., 81
Proto-Austroasiatic, 457
Proto-Basque, vi, viii, xxiii, 6, 8, 9, 93, 95, 119, 122, 123, 125, 127, 137, 140, 141, 143-145, 148, 159, 170, 173, 174, 181, 182, 184, 194, 232, 233, 238, 240, 242, 246, 260, 264, 268, 271, 296, 321, 323, 326, 356, 360, 371, 372, 395-397, 422, 424, 454, 455, 457, 458, 460
  Michelenian Proto-Basque, 125, 137, 162, 163
  non-Proto-Basque, 218
  Old Proto-Basque, 9, 119, 124, 125, 137
  Pre-Proto-Basque, 6, 9, 321, 327, 391
Proto-Vasconic, *see* Old European
Ptolemy, 70-72, 75, 83, 113, 115
Pyrénéen(s) occidental/-aux ancien(s), Language, roots, words, 184
Pyrenees, Mountain Range of the, 11, 12, 59, 60, 73, 74, 81, 83, 92-94, 107, 116
  Pyrenean Aragonese, 168
  Pyrenean dialects (of Basque), 99, 108
  Pyrenean stratum, 178
Pyrénées-Atlantiques, Département des, 11

Q

Quechua / Quechuan languages, 28, 32, 84

R

Rebuschi, G., 3, 310, 313, 314, 317, 422
Reformation, 101
Refranes y Sentencias, xxi, 100, 108, 127, 402
Reguero-Ugarte, U., 8, 97, 429
Rey, A., 186
Ribagorza, High, 75
Rijk, R. de, 3, 233, 240, 244, 246, 249, 250, 257-259, 261, 270, 277-279, 329, 337, 359, 378, 390
  Basque as an SOV language, 443, 444
  directional all., 267
  NRel structures, 452
  origin of the ines., 250-252, 254, 255
  origin of the verbal *da-*, 260, 375
  terminative all., 268, 269
Rijkhoff, J., 458
Ringe, D., 178
Rioja, La, 70, 85, 94, 116, 151
  High Rioja, 45, 85
  Lower Rioja, 115
Ripalda, G., xviii
Rodríguez-Adrados, F., 73
Rollo, W., 288
Romance
  R. influence, 2, 3, 64, 91, 95, 96, 104, 110, 124, 134, 145, 148-150, 152-154, 156,

158, 162, 175, 176, 219, 220, 228, 229, 294, 329, 331, 332, 334, 345, 354
R. languages, xxiv, 34, 61, 90, 95, 100, 115, 176, 354
R. linguistic phenomena, 86, 96, 124, 149, 176, 216, 217, 265-267, 277, 289, 444
R. linguistic tradition, 2, 4
Roncal, Valley of, xxii, xxiii, 56, 58, 60, 68, 77, 78, 116, 135, 136, 401
Town, 60
Roncalese (subdialect of S), xxii, 8, 59, 60, 67, 80, 102, 108, 114, 125-127, 131, 134, 137, 139-141, 149, 163, 164, 168, 169, 177, 221, 233, 258, 285, 290, 325, 340, 408
independent dialect, 68, 69
Ross, M., 178
Rotaetxe, K., 225
Ruiz-Arzalluz, I., 186, 360

S

Sabando, xix
Sagart, L., 180, 196
Saint-Jean de Sorde, Cartulary of, 97
Sainz, K., 189, 195, 239, 325, 326, 360, 391, 404, 430, 456, 458
Sakana, 8, 41, 46, 49, 158
Middle Sakana, xxi, 47, 49, 304
Salaberri, J. D. J., 109
Salaberri-Muñoa, P., 103
Salaberri-Zaratiegi, P., 46, 86, 97, 116, 223
Salaburu, P., 44, 49, 293, 304
Salazar, B., 288
Salazar, Valley of, xxii, xxiii, 56, 57, 68, 135, 136, 401
Variety of, 57
Salazarese (subdialect of East.-LN), xxiii, 8, 57, 59, 67, 80, 108, 168, 169, 232, 249, 259, 285, 290, 325, 408
Salmons, J. C., 175
San Juan de la Peña, Monastery of, 97
San Millán
Cartulary of, 143, 169
cult of, 85
Monastery of, 45, 96, 97
Reja de, 45, 85, 143
Santa Grazi, 60, 78
Santazilia, E., 7, 223, 230, 231, 247
Sapir, E., 445
Sara, xvii, xix, 8, 103, 108, 339
Variety of, 52
Saramaccan creole language, 196
Sarasola, I., xxv, 90, 99, 112, 197, 304, 313, 314, 316
Satrústegui, J. M., 102, 286
Sauguis, B. de, 100, 128
Saussure, F. de, 89, 179
School of Sara, 53, 103
Schuchardt, H., 184, 200, 227, 254, 257, 263, 269, 360, 363-365, 368, 372, 388-390, 392, 393, 400, 425
dat. flags, 329, 386, 387, 392, 393
dativus ethicus, 397, 398, 402, 404, 405, 409
figure of Bascology, 2
origin of *-eta*, 236, 238, 289
passivity of transitive verbs, 383, 394, 422
soc. *-(ar)e-kin*, 269, 270
Schuh, R. G., 195
Second World War, 110
Semitic languages / language family, 115, 173, 179, 180, 205
Serial Verb Constructions, xxiv, 4, 195, 206, 229, 261, 269, 271, 391, 416, 424, 426, 427
Siewierska, A., 4, 458
Sinitic languages / language family, 173, 205
Sinosphere, Languages of the, 191
Slavic languages / linguistic tradition, 2
Smith, J., 171
Sohüta, xviii, 78
Solnit, D. B., 196
Soroa, M., 110
Southeast Asian languages, 417
Southern Guipuscoan (subdialect of G), 43
Spanish Civil War, 111
Spanish language
influence of the, 24, 127, 302, 332, 382, 383
linguistic phenomena of the, 242, 301, 302, 345, 355, 368, 381, 397, 408, 412, 416, 444
Spanish Second Republic, 16
split ergativity, 377, 423, 424
St. James' Way, 76
Stampe, D., 178, 191, 198, 201, 204, 457
Standard Basque, xx, xxiii, 8, 9, 35, 39, 53, 120-122, 224, 225, 227, 284, 323-325, 333, 335, 337, 342, 347, 442, 447
Stassen, L., 190, 271
Stempf, V., 394

Strabo, 71, 115
Sturtevant, E. H., 73
subject-predicate discourse structure, 459
subject-prominent languages, 395

T

Tamazigh language, 28
Tani languages / language family, 192, 203
Tartas, J. de, xviii, xxi, 102, 103, 128, 343
Tejerina, B., 16
Thompson, S. A., 438, 459
Tibeto-Burman languages / language family, 180, 196
TOG phenomenon, *see* Transitive Objective Genitive
Tolosa, 43, 77, 266
  Variety of, 43
topic-comment discourse structure, 459
topic-prominent languages, 395
Torreira, F., 166, 168
Tovar, A., 5, 73, 184, 186, 188, 214
Transitive Objective Genitive, 66, 417
Trask, R. L., vii, xxv, 2, 3, 34, 81, 189, 190, 283, 285, 326, 446
  figure of Bascology, 3, 4, 8, 9
  genetic relationships, 1, 174, 183
  noun declension, 65, 223, 225, 227, 228, 236, 238-241, 245, 247-250, 252, 253, 256, 259, 264, 266, 268, 270, 273, 277, 291
  old consonant system, 145, 175, 176
  passivity of PB, 242, 365, 366, 383, 394, 395, 422, 423
  Trask's Law, 64, 79, 238, 265, 371, 386
  verb morphology, 193, 196, 260, 323, 324, 326-329, 331, 333, 337, 339, 341, 345, 354, 359, 360, 368, 375, 387, 392, 393
  word order, 189, 239, 430, 455, 456, 458, 459
Trask's Law, 79
Trombetti, A., 184
Turkic languages / language family, 189
Txillardegi (pseudonym of J. L. Álvarez-Enparantza), xxv, 112, 257, 450
Txorierri, Valley of, 302
Txulapain, Valley of, 294

U

Ubillos, J. A. de, xxi
Udolph, J., 115
Ugao-Miravalles, 288, 290
Uharte-Arakil, Township of
  Variety of, 47
Uhlenbeck, C. C., 73, 128, 177-179, 184, 205, 240, 263
Ulibarri, J. P., 107
Ulibarri, K., 6, 39, 47, 64, 89, 104, 114, 132, 451
Ultzama (subdialect of North.-HN), 47, 49
  Valley of, 32, 47-49, 52, 233, 293, 294, 304
Unzueta, M., 81, 84-86
Uralic languages / language family, 173, 180, 182, 205, 225
  Proto-Uralic, 189
Urbasa, Mountain Range of, xxi, 46, 49
Urdazubi, xvii, 52, 69, 293
Urdiain, 43, 49
Urgell, B., v, vii, 77, 89, 103, 104, 109, 236, 283, 292, 311, 320, 323, 327, 329, 332, 337, 339-341, 353
Uriarte, J. A., 108, 427
Uribe, Area of South, 114
Urketa
  Variety of, 57
Urola, Valley of, 39
Urquijo, J. de, xxv, 73, 107, 108, 117, 128
Urquizu, P., 99, 102, 266
Urraulgoiti, Valley of, 56
Urruña, 33
Urteaga, M. de, xviii
Urthe, P. d', xxi, 105
Urzainki, 60
  Variety of, xxii, 60
Uterga, xvii, 51
Uztaritze, 56
  Variety of, 55
Uztarroze, xxii, 60, 169
  Variety of, xxii, 60

V

Valpuesta, Monastery of, 96
Varduli, 70-72, 81, 83, 84, 113
Vascones, 70-72, 81-84, 86, 113, 115
Vennemann, T., 115, 201, 445, 457
Vetones, 84
Videgain, X., 304
Viellard, J., 127
Villabona, xxi

Villar, F., 81, 82
Villaro-Areatza, xvii
Villasante, L., 99, 110, 112
Vinson, J., 107, 117, 236, 390
Vitoria-Gasteiz, xx, xxii, 35, 45, 46, 72, 76, 77, 85, 100, 116, 341
Viva Jesus, xxi, 103, 114, 132, 155
Voltoire, xxi, 103, 105, 390

W

Watkins, C., 456
Western Biscayan (subdialect of B), 35
Western Low Navarrese from Labourd (subdialect of West.-LN), 55
Western Southern High Navarrese (subdialect of South.-HN), 50
Whitman, J., 201

Y

Yana language, 395
Yrizar, P., 4, 34, 43, 44, 52, 55, 68, 69, 114

Z

Zabala, I., 230
Zabaltza, X., 109
Zaldubi (pseudonym of G. Adema), 107
Zarabozo, C., 129
Zarautz, xvii, xxi, 106, 259
Zavala, A., 109
Zavala, Fray J. M. de, xxi, 107, 108, 409
Zavala, R., 196, 366
Zeberio, 134
Zegama, xviii, 44, 84
  Variety of, 43
Zelaieta, E., 34, 47, 48
Ziburu, xvii, xviii, xix
Zizur, Township of
  Variety of, 50
Zuazo, K., 34, 39-43, 45, 46, 49, 53, 56, 61, 62, 65, 70-72, 79, 106, 187, 202, 369, 370
  dialectal branching, 75-78, 80
  dialectal classification, 69
Zubia, N., 103, 266
Zugarramurdi, 52, 56, 69
Zumarraga, Fray Juan de, 99, 438

Mikroglottika
Minority Language Studies

Edited by Raúl Sánchez Prieto, Daniel Veith and Mikel Martínez Areta

Vol. 1 Raúl Sánchez Prieto / Daniel Veith / Mikel Martínez Areta (eds.): Mikroglottika Yearbook 2008. 2008.

Vol. 2 Raúl Sánchez Prieto / Daniel Veith / Mikel Martínez Areta (eds.): Mikroglottika Yearbook 2009. 2010.

Vol. 3 Abel Pardo Fernández: Italiano-leonese: lingue, marketing, nuove tecnologie e didattiche. 2011.

Vol. 4 Peter Gilles / Melanie Wagner (Hrsg.): Linguistische und soziolinguistische Bausteine der Luxemburgistik. 2011.

Vol. 5 Mikel Martínez Areta (ed.): Basque and Proto-Basque. Language-Internal and Typological Approaches to Linguistic Reconstruction. 2013.

www.peterlang.de